An epidemiological study of Labrador Retrievers

Carys Pugh

Thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

2016

Declaration

I declare that the research within this thesis is my own work and any assistance received has been duly acknowledged. The work described has not been submitted for any other degree or professional qualification.

Carys A Pugh

Edinburgh, January 2016

Acknowledgements

This is the hardest part of my thesis to write because I want to find the right words to thank many people. I'll start with the BBSRC, which funded my PhD. I am very grateful for the opportunity. Also, thank you to the Dogslife owners who made the project possible.

Within Roslin there are my supervisors who all listened to me fulminate about the database - I appreciate your tolerance. In strict alphabetical order: Watching Dylan deal with people when we were all down in Birmingham was a priceless. More seriously, I don't think I've known anyone quite so tenacious. He was always available and supportive, even in the face of my enthusiasm for validation and I particularly need to thank him for his patience. Ian catered to my inner nerd and kept having smart ideas. He also listened when I needed an ear and that was much appreciated. Kim was an invaluable arbiter and sounding board. She patiently explained genetics... repeatedly. She was also kind this last Summer. It could have been very difficult but she made it easier. Mark creates an atmosphere in EERA which is one of the main reasons I've enjoyed my PhD so much. I've developed a fondness for heatmaps and, when I watch him think, I realise how much I still have to learn. You've all been tolerant and supportive - thank you.

Beyond my supervisors, there are two people who have been crucial: Erica Rose and Damon Querry. I think the number of times Erica is mentioned throughout this thesis is a good indication of her indispensable contribution. More broadly, she always had a smile when I appeared at her office door. Damon didn't have a smile. He always rolled his eyes and then gave me his time anyway. Apologies for interrupting so often! You were both brilliant - thank you.

There are many people in the building who helped me with coding and analysis. I shamelessly bugged Siben with coding queries, Paul was a map guru and Gustaf helped unpick some awful error messages. Helen kindly reassured me

that my stats approach was legitimate and Lindsay and Enrique helped explain genetics to me… repeatedly.

There have been multiple generations of lunch and coffee people who all heard about the database. John, Dave, Fiona, Kate, Graham, Brendan, Andrew, Debby, Yan, Enrique, Isobel and Konrad - thanks guys. Much like the EERA group's working style, you made my time here.

To the Bronsvoort and Summers lab groups - thank you. Conferences were better with Amy and Stella and Ailsa kept me safely out of the lab.

More broadly, Nicola, I'm glad we met. I appreciate you not laughing too much at my public engagement efforts and am a massive fan of tea on hill-tops. Particularly if I haven't had to carry it.

Outside work there are more people to thank. John, I will never go back to that pub. Fiona, five is the magic number! Ashley, thank you for bringing glitter into my life.

Gemma, you've been so important as a positive, supportive force. Alex, I know Glasgow is dark, cold and wet but I can't help but be delighted that you're there. Lucy, you have a wonderful sense of fun and were extraordinarily caring during a difficult time.

Last but by no means least, there's my family. They are used to waiting for me so I suspect they won't be offended. Chris, I know how much this PhD means to you and I am starting to realise how much it means to me. I've worked hard to get here but I wouldn't have made it without your support. Thank you - this is for you. Mum, Dad, I know I always have your unequivocal love and support and that's a touch-stone in my life. I can't imagine what it would have been like without you there backing me up. Rob, Eilidh, the same goes for you two. I am lucky to have a genuinely kind brother and you both gave me a nephew and niece. Now that this is nearly over, I will be visiting more often.

Lay Summary

Dogslife is the first large-scale study that aims to follow a group of dogs throughout their lives. Owners of pedigree Labrador Retrievers across the UK have been recruited to repeatedly answer an online questionnaire about their dog's lifestyle, height, weight and health. This thesis relates to over 4,000 dogs using data collected until they were up to four years of age.

Initial work was undertaken to check that the questionnaire was well understood by the owners and that the data being collected accurately reflected the dogs' lives. Owners were visited and veterinary records collected. Both investigations yielded positive results but it appeared that owners forgot to report illnesses if they did not complete the Dogslife questionnaire frequently.

Accelerometers (devices that measure activity) were sent to a sample of the group and the results were compared to the Dogslife exercise questionnaire. The questionnaire answers were related to accelerometer readings which indicated that the dogs were being sedentary, and undertaking light and moderate to vigorous exercise.

On average, Dogslife dogs were exercised for over two hours each day with the time spent being dominated by time off lead and on other activities. Dogs in England spent less time exercising than those in Scotland and Wales, and dogs in family households spent less time exercising than those in single adult households or households comprising more than one adult. Despite being pedigree animals, the males in the cohort were 2-3cm taller than the breed standard. On average, the females met the breed standard but there was wide variation for both sexes. Working dogs in the cohort were over 2kg lighter than household pets and chocolate coloured dogs were 1.4kg heavier than their black and yellow counterparts. Dogs in multi-dog households were 0.5kg lighter than those in households with no other dog. Heavier dogs spent less time fetching, chasing and retrieving and on other exercise.

Over 6,000 signs of illness were reported to Dogslife in the first three and half years and approximately half of them did not involve a veterinary visit. Reported signs were dominated by vomiting and diarrhoea, both of which peaked when the dogs were between 3-6 months of age. For the first time, rates of diarrhoea were shown to be positively associated with human population density in the UK. Limber tail (a condition characterised by a limp tail) was found to be associated with swimming in the cohort and working dogs were more likely to develop the condition than pets. Analyses suggested that the illness might be partially due to a genetic predisposition.

Data from the Dogslife project provide a unique resource for investigating the health of Labrador Retrievers. This thesis creates a platform for all such future investigations.

Abstract

Dogslife is the first large-scale, longitudinal cohort study of canine lifestyle, morphology and health. The project involves recruiting the owners of UK-based, Kennel Club registered Labrador Retrievers and asking them to submit data about their dogs via an online questionnaire repeatedly as the dogs age. In this thesis, I have analysed Dogslife data regarding the lifestyle, morphology and health of Labrador Retrievers up to four years of age.

A validation study was initially undertaken in order to understand the quality of the Dogslife data because this would underlie all future investigations. Owners were visited and veterinary records scrutinised. It was determined that Dogslife illness reports were subject to recall decay and that minor changes would improve the usability of the questionnaire. Accelerometers were subsequently sent to a subset of the cohort and aspects of the Dogslife exercise questionnaire were found to be correlated to accelerometer readings indicative of sedentary, light and moderate to vigorous exercise.

Overall, Dogslife dogs were exercised for over two hours each day with the time spent being dominated by time off lead and on other activities. Dogs in England spent less time exercising than those in Scotland and Wales and dogs in family households spent less time exercising than those in single adult households or households comprising more than one adult. Despite being pedigree animals, the males in the cohort were 2-3cm taller than the breed standard. On average, the females met the breed standard but there was wide variation for both sexes. Working dogs in the cohort were over 2kg lighter than household pets and chocolate coloured dogs were 1.4kg heavier than their black and yellow counterparts. Dogs in multi-dog households were 0.5kg lighter than those in households with no other dog. Heavier dogs spent less time fetching, chasing and retrieving and on other exercise.

Over 6,000 signs of illness were reported to Dogslife in the first three and half years and approximately half of them did not involve a veterinary visit. Reported signs were dominated by vomiting and diarrhoea, both of which peaked when the dogs were between 3-6 months of age. For the first time, rates of diarrhoea were shown to be positively associated with human population density in the UK. Limber tail was found to be associated with swimming in the cohort and working dogs were more likely to develop the condition than pets. Genetic analyses identified regions of interest that might predispose the dogs to limber tail on chromosomes 6 and 30.

Data from the Dogslife project provide a unique resource for investigating the epidemiology of Labrador Retrievers. This thesis creates a platform for all such future investigations.

Publications

Papers

- **CA Pugh**, BM de C Bronsvoort, IG Handel, KM Summers, DN Clements (2015) Dogslife: A cohort study of Labrador Retrievers in the UK. *Preventative veterinary medicine* (in press).

- **CA Pugh**, KM Summers, IG Handel, BM de C Bronsvoort, DN Clements (2015) Validity of Internet-Based Longitudinal Study Data: The Elephant in the Virtual Room. *Journal of Medical Internet Research*, 17(4):e96.

- **CA Pugh**, BM de C Bronsvoort, IG Handel, KM Summers, DN Clements (2014) What can cohort studies in the dog tell us? *Canine Genetics and Epidemiology*, 1(1):5.

- DN Clements, IG Handel, E Rose, D Querry, **CA Pugh**, WE Ollier, KL Morgan, LJ Kennedy, J Sampson, KM Summmers, BM de C Bronsvoort (2013) Dogslife: A web-based longitudinal study of Labrador Retriever Health in the UK. *BMC veterinary research*, 9:13.

Conference Proceedings

- **CA Pugh**, BM de C Bronsvoort, IG Handel, KM Summers, DN Clements (2015) Dogslife: A cohort study of Labrador Retriever Health. *Society for Veterinary Epidemiology and Preventative Medicine*, Ghent.

Acronyms

AIC Akaike Information Criterion

AKC American Kennel Club

ALSPAC Avon Longitudinal Study of Parents and Children

CI Confidence Interval

cm centimeters

CP Carys Pugh

DAQ Damon Querry

DEID Data Entry identifier

DHR Dogslife Health Report

DNC Dylan Clements

ER Erica Rose

g grams

GDBA Guide Dogs for the Blind Association

GP General Practitioner

GPS Global Positioning System

GR Golden Retriever

GWAS Genome Wide Association Study

HRT Hormone Replacement Therapy

ID identifier

IID Illness identifier

IQR Inter-quartile range

IT Information Technology

KCID Kennel Club identifier

kg kilograms

lb pounds

LR Labrador Retriever

NHS National Health Service

NI Northern Ireland

oz ounces

PCR Polymerase Chain Reaction

RAO Recurrent Airway Obstruction

sd standard deviation

se standard error

SNP Sinle Nucleotide Polymorphism

UK United Kingdom

UKKC United Kingdom Kennel Club

USA United States of America

Contents

Chapter 1

Introduction

1.1 Introduction

This thesis comprises an epidemiological study of Labrador Retrievers (LRs). The findings reported are based on the Dogslife project (Dogslife [2015]), which collects data from the owners of a cohort of pedigree dogs in the United Kingdom (UK). As part of my studentship I reviewed canine cohort studies (Pugh et al. [2014]) and present this review below as an introduction to the thesis.

Understanding the factors relating to disease in a population is important for anticipating and dealing with health care needs. Studying the health of populations and identifying population-wide strategies to improve health can be undertaken in a number of ways. Beyond descriptive approaches, analytical studies can be split into experimental and observational investigations. Dohoo et al. [2010a] distinguished observational studies from experimental studies, where investigators control the allocation of subjects to study groups, by suggesting that in observational studies, investigators "try not to influence the natural course of events for the study subject".

Epidemiologists traditionally divide observational studies into case-control, cross-sectional or cohort study designs (Thrusfield [2005]; Dohoo et al. [2010a]). The advantages and disadvantages of each of these study types, particularly with regard to susceptibility to bias, are fully described in Table 1.1. In brief, case-control studies are particularly useful for rare diseases but lack an ability to clarify temporal relationships between events and exposures. Cross-sectional studies can be performed at a single time point and allow investigators to seek associations between potential risk factors and outcomes, but again do not allow the assessment of temporal dependencies. Cohort studies, where individuals are tracked through time, solve this problem as investigators can assess whether risk factor exposures are followed by outcomes in individuals. This element of time dependency is crucial to infer causation between risk factors and disease, and to understand transmission dynamics of infectious diseases. Further, cohort studies lend themselves to analysis of the effect of long-term exposure to a risk factor or treatment and, with targeted recruitment, are ideally suited to examine the effect of rare risk factors. Unfortunately cohort studies necessarily involve a large investment of time and finances, both to set up and maintain. Historically, they take time to yield results and have therefore been used sparingly in the field of canine disease.

Studies from human medical literature will be discussed before describing the types of canine cohort studies reported to date. Dogslife is an attempt to apply approaches more normally found in human medicine. Applied to canine epidemiology, the techniques have immense potential for health advances.

1.2 The benefits of cohort studies: Comparative examples

One of the most widely renowned cohort studies in human medicine is the Framingham Heart Study. Researchers recruited a group of over 5,000 women and men aged between 30 and 62 years old living in Framingham, Massachusetts in 1948. The cohort were evaluated every two years regarding their medical status and lifestyles, including physical examinations and collection of biological samples for laboratory testing. The study identified many of the major cardiovascular disease risk factors which we take for granted today, such as high blood pressure, high blood cholesterol, smoking, obesity and diabetes (Mahmood et al. [2013]). The analysis of the Framingham cohort has resulted in over 2,000 peer reviewed publications, and aptly demonstrates how the detailed, repeated evaluation of modestly sized cohort groups can result in the identification of risk factors for disease which have global significance.

Another important early cohort study of human health was undertaken in the UK in 1951. The aim was to address concerns about an observed association

Table 1.1: Advantages and disadvantages of different observational study types

Potential goals	Advantages	Disadvantages
Cross-sectional		
Population prevalence of exposure and/or outcome. Associations between exposures and outcomes.	Relatively simple. Relatively cheap. Relatively quick. Good for common conditions and exposures. May assess multiple exposures and outcomes. Good for initial assessment of an exposure or outcome.	Poor for rarer exposures and outcomes. No causality may be inferred as exposures and outcomes are measured contemporaneously. Highly susceptible to information bias.
Case-control		
Associations between exposures and outcome. Strength of association in the form of odds ratio between exposure(s) in controls and exposure(s) in cases.	Relatively cheap. Relatively quick. May assess long latent periods. Good for rarer outcomes.	Choice of controls notoriously difficult. May only examine one outcome. Odds ratio not an intuitive measure. Highly susceptible to selection and information bias and population stratification.
Cohort		
Incidence rates. Temporal associations between exposures and outcomes.	Good for rare exposures. May examine multiple exposures and outcomes. May assess long latent periods. May assess temporal relationship between exposure and outcome inferring causality.	Not simple. Not cheap. Not quick (unless retrospective). Highly susceptible to retention bias. Susceptible to sampling bias.

between smoking and lung cancer. To examine the question of causality, the study was designed to determine whether it was possible to predict someone's risk of developing lung cancer from their smoking habits earlier in life (Doll [1999]). Over 40,000 doctors were recruited, which was over two thirds of the doctors on the British Medical Register at the time. The study went on to investigate the impact of smoking on diseases beyond lung cancer, including vascular disease and other neoplasias (Doll et al. [2004]). Ultimately the cohort was so valuable that the members were followed for their lifetime and the last questionnaire was sent out some 50 years later.

Two more recent studies which have illustrated the power of large scale cohort studies are the Avon Longitudinal Study of Parents and Children (ALSPAC) (Golding [1990]) and the Italian NINFEA cohort (Richiardi et al. [2007]). Both are birth cohorts, initially designed without specific hypotheses in mind. Instead they set out to collect information on a variety of exposures to broadly investigate pregnancy and the early life of children. In the case of ALSPAC, investigations went back even earlier, with assessment of antenatal risks, such as the impact of maternal drinking prior to conception and in early pregnancy on birth weight (Passaro et al. [1996]).

The ALSPAC study team faced great difficulty obtaining funding in the initial years of the project (Overy et al. [2012]). As time passed and significant risk factors started to be found and reported, it became more widely recognised that the cohort was an incredible resource that should be maintained in the long-term. The cohort and findings generated by working with its members have been summarised by Fraser et al. [2013]. The open-ended investigative approach resulted in the identification of a range of phenotypes and influencing factors that could not have been predicted by the investigators at the start of the study. The costs of recruiting the cohort would have been wasted if contact with members were lost before these discoveries could be made.

Analyses of the ALSPAC cohort did not stop with exploration of early-life influences. As the costs of collecting, archiving and analysing DNA reduced it became possible to add genetic data to the wealth of phenotypic data and explore the interaction of genotype with other variables. Over more than 20 years the ALSPAC team moved from having a relationship with pregnant mothers to having a relationship with the children from those pregnancies. These children have grown to start their own families and the next generation are also being recruited into the study. A wealth of discoveries guiding national public health policy have been made during the study. These include understanding the influence of sleeping position on the risk of cot death (Golding et al. [1992]; Hunt et al. [1997]) and the benefits of eating oily fish on children's mental development (Daniels et al. [2004]), both of which have directly led to the development of guidelines for best practice.

Between 1996 and 2001, the Million Women Study recruited women over 50 in the UK (The Million Women Collaborative Study Group [1999]). Recruitment through breast cancer screening centres built-in a reliable method of ascertaining the primary outcome of interest - the incidence of breast cancer. Environmental influences were captured in a lifestyle questionnaire that was completed at recruitment and periodically thereafter. Information regarding other disease events such as incidence of fractures was also collected via the follow-up questionnaires (Banks et al. [2004]).

The main finding from the Million Women Study regarding the impact of Hormone Replacement Therapy (HRT) on the incidence of breast cancer (The Million Women Collaborative Study Group [2003]) remains controversial. An increased incidence of breast cancer was found in the women taking HRT but it has subsequently been argued that these women were more likely to be tested for breast cancer, resulting in increased diagnoses. The women involved were not randomly assigned to receive HRT so the potential for confounding cannot be ignored. Nevertheless, the study built on results from earlier cross-sectional studies and it had enormous power to detect associations. As the women were followed with time, causal inference is possible. At the very least the results of the many publications about the cohort will influence the direction of future randomised controlled trials to try and definitively determine causal relationships.

The value of cohorts has been recognised and data collected previously are increasingly the foundation for further analysis. For example, a team from Edinburgh University took advantage of historic data collection to develop a cohort of people with results that span over 80 years. In Scotland 95% of children born in 1921 and 1936 were given an intelligence test at the age of 11. The team recruited a subset of the survivors from these tested cohorts some 60-70 years later to investigate their cognitive function (Deary et al. [2012]) and the environmental and genetic influences upon them. Their continuing assessment of cognitive function has led to the discovery of an association between carrying the APOE E4 allele (also associated with Alzheimer's disease) and non-pathological cognitive decline (Schiepers et al. [2012]). The cohort is a unique resource for the investigation of the effects of ageing on cognition and it continues as participants enter their tenth decade.

While the benefits of cohort studies are well understood (Table 1.1), the extended time to finding results and relatively high costs are undeniable. In part to address these costs, the US Department of Defense started to move cohort studies into the internet age when they set up the Millennium Cohort Study of US military personnel (Gray et al. [2002]). Current and ex-military personnel were recruited and offered the chance to answer the questionnaire by post or online. The financial savings associated with participants replying online were such that they offered a $5 incentive and still estimated their savings per online response at $50 compared to those responding by post (Smith et al. [2007]). As internet access has increased, epidemiological studies have gradually made greater use of the technology. The NINFEA cohort is based entirely online (NINFEA [2014]). Whilst the costs of setting up and maintaining functional and appealing web portals are not insignificant, studies are now possible that would not have been feasible if based on face-to-face, telephone or postal questionnaires. Building on this experience of human studies, canine cohort studies that would have been inconceivable are now financially viable and the potential to exploit this avenue of research is immense.

1.3 Canine cohort studies

Despite the extensive number of findings uncovered by human cohort studies, the design has not been widely used in canine research in the past. As discussed, the cost and time burdens can be prohibitively high. A number of canine cohort studies have been reported and in each case attempts have been made to overcome the associated financial burden. The different strategies used are discussed below.

1.3.1 Retrospective methods

Retrospective cohort studies involve looking back at individuals after the events of interest have occurred (for example disease incidence, death or pregnancy) and the follow-up period has ended (Dohoo et al. [2010b]). These studies can be undertaken on a large scale with relatively little lead time or up-front costs by using pre-existing databases such as those maintained by insurance companies and groups of secondary veterinary hospitals or primary clinics. The advantages and disadvantages of this type of study are summarised in Table 1.2.

Insurance databases in particular are an extremely valuable resource and are discussed in detail by O'Neill et al. [2014a]. There is a long tradition of pet insurance in Sweden and Agria insure approximately 40% of Swedish dogs (Bonnett and Egenvall [2010]). Their database provides a powerful measure of events in the Swedish pet canine population (Egenvall et al. [1998]). Such large electronic resources offer the chance to study incidence rates and survival time from diagnosis for specific diseases, such as mammary tumours (Egenvall et al. [2005]). However there is no requirement for private companies to make their data available. When using insurance databases, there is likely to be bias relating to the non-random socioeconomic status of owners who insure their pets and to specific insurance policy exclusions such as pre-existing conditions and age limits. In addition, in countries where dog insurance rates are low, the resource would be even less representative of the population as a whole.

Veterinary medical databases provide an alternative resource of information on the health of populations (O'Neill et al. [2014a]). They have the advantage of that they can be linked to ancillary resources (such as radiographic archives and biological samples). However, the plethora of recording systems, and lack of agreement of diagnostic criteria for the definition of specific diseases, makes them cumbersome to use and extracting and extrapolating data is difficult. With modern textural mining tools there is scope to revisit this area (O'Neill et al. [2014a]) but the challenge of collating records from diverse recording systems remains. Further, when these databases rely solely on groups of specialist hospitals, there is the risk of referral bias as demonstrated by Bartlett et al. [2010].

Risk factor studies using both insurance and veterinary medical databases are also limited by the type of data collected. In both cases, the data refer to phenotype of the dog but not their wider environment. Postcode (location) data have been used to assess the spatial distribution of atopic dermatitis (Nødtvedt et al. [2007]) but the impact of the dogs' lifestyles is not available from such records. For example Glickman et al. [2011] were able to investigate a link between severity of periodontal disease in dogs and subsequent chronic azotemic kidney disease (kidney disease causing high levels of blood urea and creatinine) because

Table 1.2: Advantages and disadvantages of retrospective cohort studies

Data source(s)	Advantages	Disadvantages
Pre-existing insurance databases	Relatively cheap. Relatively quick. May assess multiple clinical exposures and outcomes. May assess long latent periods. Recruitment and retention simple.	Non-standardised diagnostic criteria. Poor generalisability in countries with high uninsured population. No requirement for insurance data to be made available. Potential for selection bias according to socioeconomic status.
Pre-existing databases from secondary veterinary hospitals	Relatively cheap. Relatively quick. May assess multiple clinical exposures and outcomes. Potential to use ancillary resources. May assess long latent periods. Good for examining serious illnesses. Recruitment and retention simple.	Non-standardised diagnostic criteria. No knowledge of wider environmental exposures. Potential for referral and geographical bias.
Pre-existing databases from primary veterinary clinics	Relatively cheap. Relatively quick. May assess multiple clinical exposures and outcomes. Recruitment simple.	Non-standardised diagnostic criteria. Non-standardised recording systems. No knowledge of wider environmental exposures. Potential for retention bias as owners move practices.

both diseases were recorded in clinical records, but environmental risk factors like diet could not be considered. This is a major limitation of such databases; otherwise their data on multiple disease outcomes, covering large numbers of dogs from different breeds, would be unparalleled in terms of potential for use in investigations.

1.3.2 Prospective Methods

Prospective studies are set up before the outcome of interest occurs and allow investigators to pre-select study subjects and specifically determine which data they wish to collect (Dohoo et al. [2010b]).

1.3.3 Prospective methods: Time-limited

Limiting the time at risk has been used to minimise the costs of studies where pre-existing data are not available. The advantages and disadvantages of this approach are summarised in Table 1.3 and one major advantage is the reduction of bias through loss to follow up. A wealth of investigations have utilised this methodology, such as those investigating the spread of Leishmaniasis and other vector borne diseases in dog cohorts. Studies investigating disease incidence (Courtenay et al. [1994]; Paranhos-Silva et al. [1998]; Moreira et al. [2003]), detection methods (Quinnell et al. [2003]; Oliva et al. [2006]; Gramiccia et al. [2010]; Otranto et al. [2010]; Quinnell et al. [2013]) and the impact of a culling regime (Moreira et al. [2004]) have all used this approach. Cohort methodology was necessary in each study but cost minimisation and swift reporting were facilitated by limiting the follow-up times to from one to three vector seasons.

1.3.4 Prospective methods: Single factor

If time is not constrained, then the focus or numbers of dogs in a study may be narrowed. The advantages and disadvantages of this type of study are summarised in Table 1.4. Perhaps the best example of this comes from a study of dietary restriction using a small group of Labrador Retrievers (48 dogs) in an experimental setting. This controlled trial has yielded an array of findings on the effect of dietary restriction on mortality (Kealy et al. [2002]; Lawler et al. [2005]), immune function (Greeley et al. [2006]), and developmental joint disease (Kealy

Table 1.3: Advantages and disadvantages of time-limited prospective cohort studies

Data source(s)	Advantages	Disadvantages
According to study protocol: May include investigators, veterinarians, breeders and owners	Costs and time according to length of the study. May assess multiple exposures and outcomes including wider environmental exposures. Good for the study of infectious diseases. Diagnostic criteria set according to study protocol. Retention bias is minimised.	Necessarily time limited so unable to assess long-term exposures and long latent periods. Recruitment not simple.

Table 1.4: Advantages and disadvantages of single-factor prospective cohort studies

Data source(s)	Advantages	Disadvantages
According to study protocol: May include investigators, veterinarians, breeders and owners	Potential to examine a single issue in great detail. May assess wider environmental exposures. Diagnostic criteria set according to study protocol.	Not quick. Potentially very expensive. Recruitment not simple. Potential for retention bias in uncontrolled conditions. May only examine multiple exposures *OR* multiple outcomes.

et al. [1992]; Kealy et al. [1997]; Kealy et al. [2000]; Powers et al. [2004]; Smith et al. [2006]; Szabo et al. [2007]; Runge et al. [2008]; Huck et al. [2009]; Smith et al. [2012]). The time-span and depth of this trial (including blood sampling and radiography at regular intervals) made it prohibitively expensive to perform on a larger scale but data on specific aspects, such as the life-long progression of osteoarthritis, could only be collected by following a cohort longitudinally in this manner.

Dobson et al. [2009] undertook a study with a similarly narrow focus but were able to recruit dogs from the normal pet population in the UK. Following 174 Flat-Coat Retrievers for up to 10 years they investigated the impact of neoplasia on mortality in that breed. Costs were also minimised in this case by contacting recruited owners just once per year for a health update and asking them to proactively contact the investigators if their dog fell ill. The study demonstrated that over 40% of the dogs died as a result of neoplasia, reducing their lifespan by three years compared to those that died from other causes.

Recruiting a large enough cohort to give the required power for an investigation and retaining that cohort to minimise bias are both key to the success of population-based cohort studies. Thrusfield et al. [1998] studied a cohort of bitches for up to five years in an attempt to assess the impact of neutering on urinary incontinence. The onus for recruiting and maintaining the cohort was placed on volunteering veterinary surgeons. Perhaps because of this responsibility, some difficulty was encountered recruiting veterinarians to participate; whilst 233 initially agreed, only 16 went on to return data (a 7% response rate). The authors made every effort to minimise bias through randomisation techniques but the potential impact of selection bias on the study should not be overlooked.

Each veterinarian was asked to recruit 40 female puppies from their practices. Should these bitches subsequently become incontinent then they were no longer followed, whilst, by design, the remaining (continent) cohort were to be followed for five years. The veterinarians received letters encouraging them to continue with the study at one and three years, and a request to contact the involved owners to check that their dogs were not incontinent after five years. The authors cite slow initial recruitment as the main reason why only 504 dogs from an original 809 were followed for the full five years. They did not directly address how many of the remaining 305 dogs were lost to follow-up (only 22 developed incontinence) but retention bias may well have affected their results. Nevertheless, by focussing on a single phenotype and spreading the responsibility for dealing with recruited animals amongst a number of veterinarians, it was possible to follow enough dogs to determine that neutered bitches had a risk of urinary incontinence that was nearly eight-fold that of intact bitches.

1.3.5 Prospective methods: Hypothesis generation

Beyond studies that focus on one disease or one exposure, there has been a movement in canine epidemiology toward the broader studies undertaken in human medicine such as the example of ALSPAC mentioned above (Overy et al. [2012]).

Table 1.5: Advantages and disadvantages of prospective cohort studies for hypothesis generation

Data source(s)	Advantages	Disadvantages
According to study protocol: Animals typically population-based but data may be generated by investigators, veterinarians, breeders and owners	May assess multiple exposures and outcomes including wider environmental exposures. Diagnostic criteria set according to study protocol. Potential to describe health and lifestyle of current population. Potential to assess the broad impact of lifestyle on disease. Potential to generate new hypotheses.	Not quick. Not cheap. Delay to results and lack of specific focus make funding difficult. Recruitment not simple. High susceptibility to retention bias. Potential for poor diagnostic accuracy if reliant on owner reporting.

The advantages and disadvantages of this type of study are summarised in Table 1.5. These studies do not necessarily aim to test a single hypothesis but rather gather data to identify new areas of investigation. In canine medicine, questionnaires have been developed that cover a wide range of potential exposures and disease outcomes and they are directed at breeders, owners and veterinarians. These studies have the disadvantage of relying on non-standardised data inputs where each animal is assessed by a different person with disparate (or no) training. However the studies are able to recruit more participants, and their subjects are more representative of dog lifestyle in the wider population than those followed under controlled conditions.

A 10-year cohort study of pedigree Boxers in the Netherlands recruited over 90% of the litters born in 14 months of 1994-5, initially comprising 2,629 puppies. The study used diary-format records and face-to-face assessment with the breeders but moved on to six-monthly questionnaires with owners. Pre-weaning mortality (Nielen et al. [1998]; van der Beek et al. [1999]) and post-weaning mortality (van Hagen et al. [2005a]) were assessed and, due to the large numbers of dogs involved in the study, all with pedigree information, the investigators were able to make heritability estimates for phenotypes (Nielen et al. [2001]) and common diseases such as cryptorchidism (failure of one or both testes to descent to the scrotum), cranial cruciate disease (degeneration of the cranial cruciate ligament) and epilepsy (Nielen et al. [2003]) (a neurological disease characterised by the development of seizures) and hip dysplasia (a developmental malformation of the hip joints) (van Hagen et al. [2005b]).

Similarly a group in Norway followed a cohort of 700 from four breeds of large dogs. Again they gave questionnaires to breeders and owners but they also involved the dogs' veterinarians. To date they have published studies on the prevalence and risk factors of neonatal mortality (Indrebø et al. [2007]), the effect of weight and growth rates on the development of hip dysplasia (Krontveit et al. [2010]) and the incidence and risk factors associated with vomiting and diarrhoea (Sævik et al. [2012]).

The Dogslife Project (Dogslife [2015]) is more recent than these efforts and is focussed on the owners of Kennel Club registered LRs in the UK (Clements et al. [2013]). It is limiting costs by utilising a website-based questionnaire and has recruited over 5,600 dogs in five years. As a prospective study, it was possible to specifically tailor the questionnaire to address areas of interest. Data collection includes detail regarding phenotype and lifestyle which will be examined with reference to dog health. Like the studies in Norway and the Netherlands, the Dogslife Project is an attempt to develop a large-scale cohort of dogs with thoroughly documented history, similar to those cohorts found in human medicine.

Since Dogslife began in 2010, a new cohort of 3,000 Golden Retrievers (GRs) has been recruited in the United States of America (USA) (Guy et al. [2015]). This study is smaller but involves veterinarians as well as owners and the reported protocol suggests that biological samples will be taken from all participating dogs.

1.3.6 The future of canine cohort research

With the relative dearth of cohort studies in canines to date, there is scope to address new questions in the future. For example, the cohort of Dutch Boxer dogs discussed earlier were reported to have a pre-weaning mortality rate over 20% (Nielen et al. [1998]). Such a loss is a clear welfare problem for dogs and more detailed studies of potential risk factors could have a great impact. Indrebø et al. [2007], Nielen et al. [1998] and van der Beek et al. [1999] each address early mortality through cohort studies but their findings focus largely on factors from birth onward. van der Beek et al. [1999] included an analysis of inbreeding coefficients but found that genetic effects in general had less effect than environmental effects at puppy and litter level. Relatively short cohort studies including the lifestyle of the dam prior to birth may shed new light on risk factors associated with both still-births and early mortality, minimising distress in owners who are currently unable to prevent early losses.

Once cohorts are in place, they are ideal for undertaking smaller, nested case-control studies. The main concern with case-control studies is the choice of controls because controls have to have come from the same population as cases. The only consistent difference between a case and a control should be the disease of interest. If cases and controls are chosen from the same cohort then the difficulty of choosing appropriate controls is easily overcome. There have been multiple examples of this technique in the ALSPAC cohort such as a study investigating the potential association between insulin-like growth factors during pregnancy and later development of cervical and breast cancers (Jeffreys et al. [2011]). This is a good example because it involved assaying blood samples from 69 breast cancer cases, 151 cervical cancer cases and 443 controls; the samples having been collected many years earlier during pregnancy. Such an investigation would have been prohibitively expensive with the whole cohort but was possible with this smaller sample.

Cohort studies such as those undertaken in human medicine could play a vital role in canine medicine if veterinarians are to be able to offer advice to owners on minimising the risks of developing disease and injury. Beyond death in very early life, morbidity and mortality in dogs in developed nations reflects the epidemiological shift in morbidity and mortality in human medicine from infectious diseases to non-communicable diseases. This shift is increasingly relevant in canine health as vaccination, antibiotics and better veterinary care ensure that more dogs in developed nations live to suffer from developmental diseases and diseases of ageing. Bonnett et al. [1997] demonstrated that whilst the highest mortality rate in dogs over six weeks of age in Sweden was trauma (typically car accidents), the next highest rate was due to tumours, followed by locomotor problems. Cohort studies of canine lifestyle have the power to investigate the risk factors associated with developing these non-communicable diseases, facilitated by the release of a draft canine genome sequence (Kirkness et al. [2003]; Lindblad-Toh et al. [2005]) and the increasing access to high density genotyping and eventually low cost whole genome sequencing. Since the dog has a shorter lifespan than humans, associations between genetic variation and disease that are also relevant to human ageing are likely to be revealed.

Human medicine is again ahead of the veterinary field with regard to incorporating biological data in cohort studies. UK Biobank has recruited 500,000 people between 40-69 years of age. The investigative team collect blood, saliva and urine samples, phenotypic data and the agreement of all participants to have their health status followed. The collection of genetic information in particular adds a new element to the traditional cohort study, and with such a large cohort the potential power to detect risk factors involving genetic-environmental interactions is enormous.

The ALSPAC team collected multiple blood, urine, hair, nails, saliva and placenta samples from the mothers and children in their study. They have used Single Nucleotide Polymorphism (SNP) chips to genotype the samples and also offer results of assays at various ages for investigation by the academic community (ALSPAC [2015]). At time of writing in September 2015, they offer cleaned Genome Wide Association Study (GWAS) data for over 8,000 parents and over 8,000 children; each individual having SNP data relating to over 500,000 SNPs. With phenotype data, that is an incredible resource and has contributed to diverse genetic studies of, for example, lung function (Repapi et al. [2010]) and body mass index (Stergiakouli et al. [2014] and Warrington et al. [2015]).

Projects on such a scale are currently financially prohibitive in dogs but, as will be discussed later, buccal swabs have been collected from a subset of the Dogslife cohort for DNA extraction, enabling comparisons of genotype with phenotype. The GR study discussed above (Section 1.3.5) will collect samples from all of the dogs in their smaller study. Should such sampling be repeated throughout the lives of the dogs, it would be possible to investigate epigenetic changes during a lifetime. The merging of lifestyle and whole genome data should increasingly reveal associations between genotype and environment in the dog and ultimately in humans as well.

1.4 Conclusions

Cohort studies have already yielded results in the field of canine health. With the advent of large databases and internet technology the costs of such studies are being reduced to the point whereby large-scale studies are possible in canine populations. The Dogslife project is the first attempt to undertake such a large-scale cohort study of canine health, genetics and lifestyle. The potential to identify risk factors and inform an evidence-based medicine approach to preventative health measures in dogs mean that Dogslife and other cohort studies such as that of the GRs can have a great impact on dog health and welfare. This thesis will report on findings from Dogslife. It will mention potential improvements that could be made for future cohort studies but largely focusses on describing the lifestyle, morphology and health of a group of pedigree LR. The findings to-date indicate just how much is possible when large groups of people work together to improve canine health.

Chapter 2

Dogslife Project & Participants

2.1 Introduction

The Dogslife project was launched in 2010 and was the first, large-scale, population-based epidemiological study of dogs. It was developed through collaboration between epidemiologists, clinicians and geneticists based at the Universities of Edinburgh, Liverpool and Manchester. I joined the study as a PhD student in October 2011 when data had already been collected for 15 months. The study was approved by the Veterinary Ethical Review Committee of the University of Edinburgh and initially funded by the Kennel Club Charitable Trust.

The aim of Dogslife was to describe dogs' lives and seek associations between their morphologies, lifestyles, genotypes and health experiences. At its inception, the intention was to recruit thousands of dog owners and ask them to give information about their dogs throughout their dogs' lives. This chapter will include an explanation of the study design, with descriptions of recruitment and retention efforts during the first three and a half years of the project. The Dogslife questionnaire will be introduced and, finally, the cohort of owners who contributed between July 2010 and 31st December 2013 will be analysed. The methodology underlying the Dogslife project has been detailed by Clements et al. [2013] and analyses of recruitment and retention for dogs up to four years have been reported by Pugh et al. [2015b].

2.2 Dogslife Study Design

As Dogslife is based in the UK, the study population is UK-based dogs. Unfortunately the UK dog population is poorly characterised. There is no register of dogs and estimates of population size differ. For example, an academic study of the UK pet population in 2007 estimated dog numbers to be 10.5 million (95% Confidence Interval (CI): 9.6 - 11.4 million)(Murray et al. [2010]) and more recently an estimate generated using unspecified methodology found a lower number of nine million (Pet Food Manufacturers Association [2014]). Dogslife was designed to address a known population; that of pedigree dogs registered with the United Kingdom Kennel Club (UKKC). Rather than include all breeds, and risk having too few dogs from some breeds to make data analysis feasible, LRs, the most popular breed of dog in the UK, were selected. This had the advantage of streamlining the study, making it more affordable financially. The study population comprised LRs born on or after January 1st 2010, living in the UK and registered with UKKC. In addition, in the interests of maximising recruitment, LR Guide Dog puppies were eligible for enrolment. These dogs were not registered with the UKKC because the Guide Dogs for the Blind Association (GDBA) maintain their own pedigree records.

Dogslife is based online and all data were collected from participating dog owners via the Dogslife website (*www.dogslife.ac.uk*). Data were recorded in a live MySQL database, maintained on University of Edinburgh servers. The live database was backed-up to a secure, offline server every six hours. The website was designed and built by an external contractor, JEM Digital. Internally, the Dogslife team included an Information Technology (IT) expert, Damon Querry (DAQ).

Offline support was available to owners from the Dogslife administrator, Erica Rose (ER). She worked full time on Dogslife and, in broad terms, was responsible for supporting owners, facilitating recruitment, maximising retention and enabling any investigations that required direct contact with the owners.

2.2.1 Recruitment Methods

Pedigree dog breeders register puppies with the UKKC after birth, generating a unique Kennel Club identifier (KCID) for each dog. The breeder may then either keep the dog or pass it on to new owners. If the puppy is passed on, the registration may be transferred by the new owner but there is no obligation to do so. The UKKC facilitated the project by notifying Dogslife of all new LR registrations. An initial bulk upload of all registrations between 1st January 2010 and 30th June 2010 was made on 1st July 2010 and then updates of new registrations were made nightly by automated electronic file transfer. These data were automatically added to the Dogslife database to give a complete list of dogs eligible to join Dogslife, uniquely identifiable using their KCID. In addition, upon request from GDBA, individual Guide Dog puppies were manually added to the database with a unique, false KCID to permit owner registration with Dogslife.

At the time of UKKC registration transfer, the new owner is able to choose whether to allow their email and postal address details to be passed onto third parties such as the study team running Dogslife; their options were 'yes', 'no' or 'never'. As part of the electronic file transfers, details of owners who transferred their dog's registration were sent to Dogslife. If the owner said 'never' then a blank line would be included, if they said 'no', all that would be included would be a name such as 'Miss A Smith'. If they said 'yes', their record would include a postal address and/or an email address according to their wishes.

Figure 2.1 is a schematic of Dogslife recruitment process taken from the paper which described the first year of Dogslife Project (©BioMed Central, Clements et al. [2013]). Following registration with the UKKC (both initial and transfers), owners were sent an information package by the UKKC. Included in this package was an A5, black and white flyer advertising Dogslife. If the owner did not register their dog with Dogslife and their email address was available, they would be sent a recruitment email. One week later, if that dog was still not registered with Dogslife and their postal address was available, the owner would be sent a coloured postcard. The only exceptions to these rules were dogs that were registered with the UKKC between January 1st and July 1st 2010, prior to the start of Dogslife. The owners of these dogs did not receive flyers or postcards. Instead, if the owner gave permission for the UKKC to share their email address, they were emailed once in early July 2010.

In addition to contacting new dog owners, there was an effort to advertise the project to LR breeders, particularly in in the period shortly after launch. In July 2010, nearly 600 breeders listed on the Accredited Breeder page of the UKKC website were contacted by email or phone. A further 1,300 breeders who

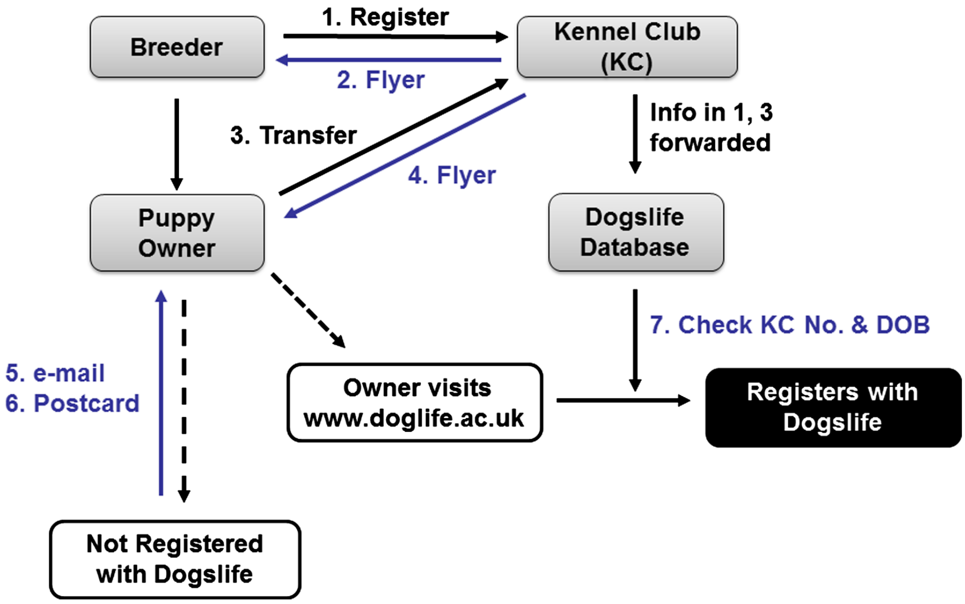


Figure 2.1: Schematic of recruitment process
(©BioMed Central, Clements et al. [2013])

advertised litters online were contacted between July and September 2010. In early 2011, 4,287 letters and flyers were sent to breeders who had had litters during 2010. Subsequently, the project administrator continued to telephone and email LR breeders who advertised their litters on the UKKC website. She asked that they encourage puppy buyers to join Dogslife.

2.2.2 Registration Process

Owners registered for Dogslife using their dog's KCID and date of birth as detailed in the screen-shots of the process in Appendix 1. The KCID and date of birth must match those sent by the UKKC, ensuring that owners were registering LRs born on, or after, 1st January 2010. During registration, owners were asked for their dog's coat colour, their own contact information (email address and telephone number) and for demographic information about their household. For both email address and telephone number, the owner would be asked whether members of Dogslife team might use the information to contact them. Owners were also given the opportunity to opt out of receiving the monthly email newsletter.

2.2.3 Questionnaire & Retention Methods

The questionnaire (correct as of May 2015) is detailed in Appendix 2. At the end of each completed questionnaire, owners were thanked and asked to return in one month's time if their dog was under one year of age or three months' time for dogs over one year. Retention was facilitated by a system of automatic reminder emails which dynamically shifted according to the last completed questionnaire for dogs under one year. Assuming permission was given to contact the owner by email and/or telephone, for dogs under one year, an automatic reminder email was sent 37 (30 + 7) days after the last data entry. This would be followed by a telephone call or, if telephone contact was not permitted or possible, non-automated email at 44 (30 + 14) days. If the owner still did not return, there would be another automatic email at 84 days and a telephone call or non-automated email at 91 days. After that, the owner would be able to login and enter data as normal but they would be considered lapsed and would not be contacted about the questionnaire. They would still receive an email on their dog's birthday and, if they had opted to receive Dogslife newsletter, they would also receive that monthly.

For dogs over one year, at the end of each data entry, owners were asked to return in three months time. However, the automated system of reminders worked on fixed three-monthly intervals when their dog was 15, 18, 21 months of age etcetera. Using the example of the 15 month (455 day) data entry, the owner would be automatically emailed at 462 (455 + 7) days provided they had not completed a questionnaire in the 21 days prior to the due date, i.e. provided there was no data entry on or after the dog was 434 (455 - 21) days old. If they did not complete the questionnaire after the 462 day reminder email, a non-automated email or a telephone call would follow at 469 (455 + 14) days. In all cases, automatic reminder emails would be sent on the exact days mentioned but non-automatic reminder emails and telephone calls would only take place during business hours Monday-Friday.

Dogslife produced a monthly newsletter (archived here: *www.dogslife.ac.uk/newsletter/archive*) including articles on dog health and behaviour, and topical issues such as advice about how to care for a dog during the autumnal firework season. It enabled the project team to update participants on Dogslife progress and facilitated recruitment of subsets of the cohort for specific studies.

Figure 2.2: Featured Dog

Every three months, a prize draw was held for Pets At Home vouchers. For the first year, there were £300 worth of prizes each quarter; 1x£100, 2x£50 and 10x£10. From September 2011, this was reduced to £200; 1x£50, 2x£25 and 10x£10 and it dropped further in September 2012 to £100; 10x£10. Owners were automatically entered into the draw each time they completed a questionnaire so the more often they visited the site and completed a questionnaire, the better chance they would have of winning. By necessity, only owners who permitted Dogslife to contact them could be eligible for the prizes and if a winner did not reply to communication, a new winner would be drawn at random. Prize winners would were also announced in the newsletters.

In addition to the various methods of maintaining contact with participants, the Dogslife website included links to dog health sites and offered a scrap-book facility. The scrap-book enabled owners to keep a photographic and written record of their dog's life for their own enjoyment. On the front page of Dogslife website, owners are also able to submit a photograph and nominate their dog as a 'Featured Dog' (for example, Figure 2.2). Submissions are reviewed for inappropriate content and dogs become Featured Dogs in order of submission. On average, a new Featured Dog is chosen three times a week. The dog's picture and owner comment is shown on the front page of Dogslife website and visitors may click through to the complete Featured Dog archive (*http://www.dogslife.ac.uk/featured/archive*).

2.2.4 Biological Sampling Methods

It was hoped that in addition to collecting data from owners about lifestyle, morphology and health, Dogslife might also investigate the dogs' biological markers and genotypes. To that end, samples of DNA and faeces were collected from a subset of the cohort. The initial aim was to bank samples for future analyses. Owners were chosen on the basis that they had answered the questionnaire at least three times and that one questionnaire was completed before the dog was six months old and another completed after the dog was one year of age; thus giving a consistent minimum period at-risk. The owners also had to have given permission for Dogslife to contact them by email or telephone. Dogs were selected in November 2012, January 2013 and April 2013 with increasing numbers meeting the eligibility criteria each time. An article was placed in Dogslife newsletter mentioning the proposed sampling which also generated volunteers. In total, 875 dogs were selected (including 18 volunteers) and owners who agreed to sampling were sent an Oragene-Animal collection tube (DNA Genotex, Ottawa, Canada) with instructions on how to collect a saliva sample. Saliva samples were collected from 479 dogs and DNA was extracted according to the manufacturers instructions and stored at -80°C. Faecal samples were collected from 383 dogs. Owners were sent a sterile 20 mL sample vial and asked to place a portion of stool into the tube. Samples were stored at -20°C prior to extraction of DNA.

2.3 Participant Profile

2.3.1 Data Cleaning & Descriptive Analysis Methods

All data described in this thesis were examined, tidied, tabulated and graphed using R (R Core Team [2013]). Specific packages used are given in *italics*.

Data cleaning was ongoing throughout Dogslife. Owners and researchers might suggest data changes when potential errors were spotted. All such amendments were made to the live database and recorded in a document that was maintained by the team (changes undertaken and recorded by DAQ or Carys Pugh (CP)). Despite these efforts, much of the data required additional cleaning prior to analysis. This final cleaning did not affect the main, live database. Instead code was written for each different data type, that exported the 'raw' data from the database using the *RMySQL* package (James and DebRoy [2012]) and output the 'tidy' data into separate files in csv format. This was true for all data cleaned prior to analysis.

Participant demographic data were collected as part of a series of one-off questions were asked during registration. These included the household type and whether (and how many) other pets there were in a household (Appendix 1). In each case, the owners were able to choose options from a drop-down list and these lists included an option of 'other' which would generate a free-text box. All free-text answers had to be examined as part of the data cleaning process in order to group together like answers.

Once cleaned, the data were examined graphically and, where appropriate, transformed. Associations were then sought using χ^2 tests and, for low numbers, Fisher's exact tests (Fisher [1922]). Where multiple comparisons were undertaken, conservative Bonferroni corrections were applied (as described by Dohoo et al. [2010c]). Where time to event data were considered, Cox proportional hazards models (Equation 2.1) (Cox [1972]) were applied using Therneau's *survival* package (Therneau [2014]).

$$\lambda(t|X) = \lambda_0(t)exp(\beta_1 X_1 + ... + \beta_n X_n) \qquad 2.1$$

Retention within the cohort was considered using a Cox proportional hazards model to seek associations between demographic factors and assumed loss to the project. For dogs under and over one year of age respectively, an owner was assumed to be lost if they had not completed a questionnaire within 60 (30 + 30) days or 120 (90 + 30) days of the 31st December.

Postcodes were grouped into countries, postcode areas and postode districts whereby EH25 9RG would be part of Scotland, EH and EH25.

2.3.2 Recruitment Results

At midnight on 31st December 2013, a copy of Dogslife database was captured. It comprised data collected over a period of three years and six months for dogs that were aged up to four years. During the period mentioned, 151,182 dogs were eligible to join Dogslife and names were passed to Dogslife for 83,532 owners who transferred their dog's registration. Contact details were included for 50% (41,476/83,532) by email and 60% (50,109/83,532) by post; 62% (52,181/83,532) by at least one method. Recruitment rates in the period until 31st December 2013 were 2.8% (4,307/151,182) of eligible dogs associated with 7.9% (4,148/52,181) of contactable owners. Recruitment rates over the first year according to different contact methods are detailed by Clements et al. [2013].

2.3.2.1 Postcode area recruitment rates

Geographically the owners were distributed across the UK and additionally on Jersey, Guernsey and the Isle of Man. A breakdown of locations is as follows: England 3227 (77.8%); Scotland 591 (14.2%); Wales 151 (3.6%); Northern Ireland (NI) 63 (1.5%); Isle of Man 9 (0.2%); Jersey 5 (0.1%); Guernsey 1 (0.02%); Insufficient information 101 (2.4%). The postcode area recruitment rates are plotted in Figure 2.3. The denominator contains information for the 50,109 eligible owners whose postcodes were available so the rates are overestimates.

2.3.2.2 Missing data

In addition to the 4,148 participating owners mentioned above, 20 additional owners registered and had unique identifiers but no associated information or dogs. These 20 owners presumably began the registration process but stopped at a very early stage.

The first aspect of owner data collected during registration was an email address. All 4,148 owners gave an email address and 3,875 gave permission for Dogslife to use that address to contact them about their dog. Each type of data subsequently collected suffered from some degree of missing or nonsense data enumerated in Table 2.1. The 62 people who gave no title or name details did not give any further information; presumably leaving the website before completing registration with the project.

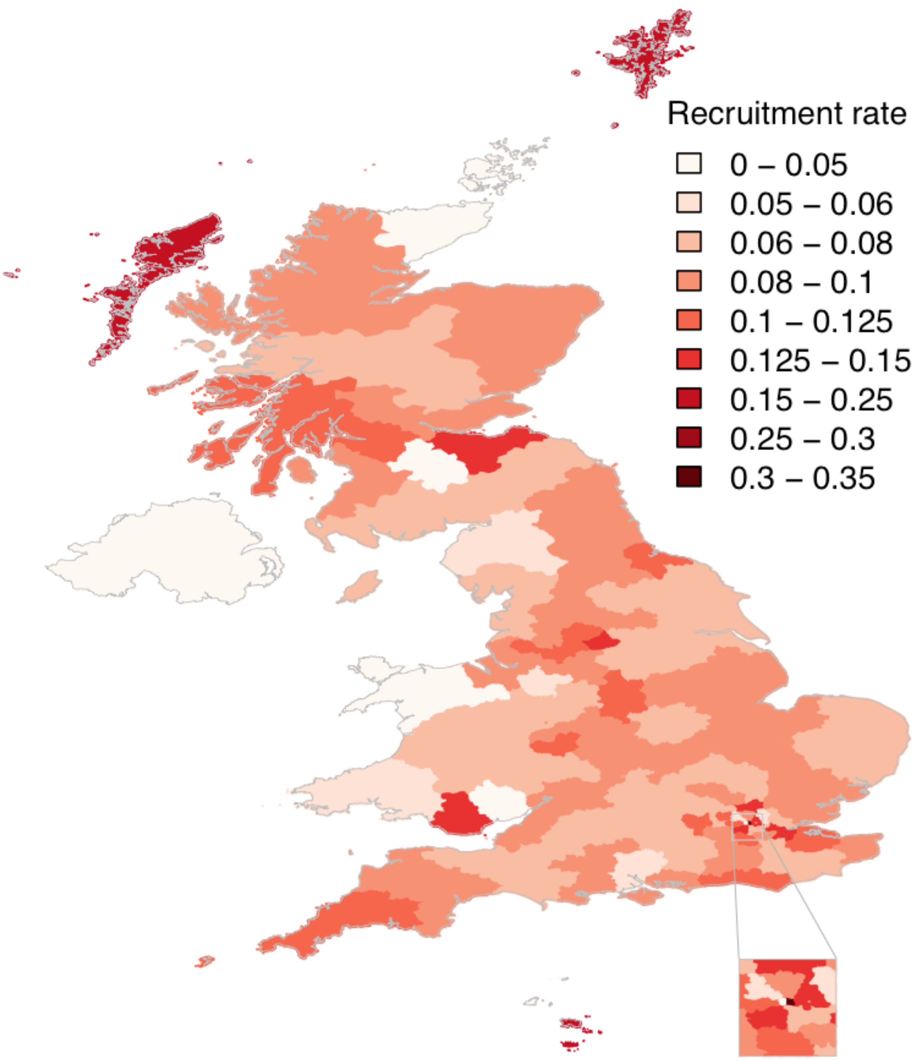


Figure 2.3: Postcode area recruitment rates showing the proportion of eligible dogs in each postcode that registered with Dogslife. Postcode data were only available for 50,109 eligible owners, just a fraction of the owners of the 151,182 eligible dogs. The denominators are therefore too low meaning the rates are overestimates.

Table 2.1: Data missing from registration

Data Type	Complete	Incomplete or nonsensical	Missing
Title	4086	0	62
Forename	4072	[a]13	62
Surname	4083	[a]2	62
Telephone Number	[b]4064	[c]19	65
Postcode	4022	[d]10	97
Household Type	4026	0	122
Smoking Status	4033	0	115

[a] For example an initial only.
[b] A series of numbers that had the correct format and number of digits. Numbers were not telephoned to check whether they were valid.
[c] For example a series of zeroes or the phrase 'ex-directory'.
[d] Five owners gave just the first half of their postcode; five gave postcodes that contained typographical errors or that did not exist.

2.3.3 Demographic Information

2.3.3.1 Household type results

There were 19 different descriptions of 'other' households reported to Dogslife and these were grouped into to the original categories according to the following rules:

1. If there were any children under 16 then the household was considered a 'family'. The corollary of this is that all reported 'families' were considered to include children under 16 years of age.

2. If someone in the household was retired then, provided there were no children under 16, the household would be considered 'retired'.

3. If all children were aged over 16 then the household was considered 'more than one adult'.

4. If left blank then the household type was categorised as 'not given'.

After cleaning, the frequencies of different household types were as follows: family 1,914 (46.1%); more than one adult 1,746 (42.1%); retired 287 (6.9%); single adult 235 (5.7%) and not given 125 (3.0%).

2.3.3.2 Other pets in household results

The final step of registration involved asking owners about other pets (if any) in their household but it was not possible to distinguish between those that did not answer the question, and those that had no other pets. Some 1,719 owners did not report any other pets. This 1,719 included 112 of the 122 who did not describe their household type and 110 of the 115 who did not give their smoking status.

The categories of other pets offered were ‘dogs’, ‘cats’ and ‘other’. There were 239 different types of ‘other’ pets reported. After cleaning, the types of other pets in households could be consolidated into 13 different categories as follows: 1,284 other dogs; 922 cats; 187 rabbits; 141 fish; 123 rodents; 115 guinea pigs; 70 birds (excluding poultry); 65 poultry; 45 horses; 31 reptiles; 9 ferrets; 6 livestock and 6 other. The number of each type of pet was often not reported so the numbers detailed are the number of owners or households with this type of animal. Each owner or household may be included multiple times.

2.3.3.3 Smoking status results

The raw reported numbers were that 735 households included a smoker and that there were 3,298 non-smoking households; that is 18.2% of the cohort. Validation efforts were undertaken and will be discussed in the next chapter. Section 3.3.1.2 indicates that there was misclassification of smoking status whereby some owners incorrectly reported that nobody in their household smoked, so the 18.2% figure may be an underestimate.

2.3.3.4 Demographic correlations

Frequencies of reported titles and household types are shown in Table 2.2. Of those titles with a clear gender definition, 76.7% were female (95% CI: 75.3 - 78.0). The five most frequent surnames were Smith (50), Jones (39), Williams (24), Wilson (24) and Taylor (21). There was no apparent preference according to household type regarding likelihood to give permission for Dogslife to contact a household by email (Table 2.3) but this changed with regard to telephone numbers. In particular, people who reported themselves to be part of retired households were more likely to give Dogslife permission to contact them by telephone ($P << 0.001$). In total, only 207 people (fewer than 5% of owners) refused

Table 2.2: Owner demographic data

	Family	More than one adult	Retired	Single adult	Not given	**Total**	%
Mrs	1073	764	173	53	26	**2089**	*50.4*
Mr	403	388	89	43	15	**938**	*22.6*
Miss	278	384	5	72	15	**754**	*18.2*
Ms	86	103	3	45	4	**241**	*5.8*
Dr	19	25	3	5	0	**52**	*1.3*
Other	3	6	0	0	0	**9**	*0.2*
Professor	0	3	0	0	0	**3**	*0.001*
Not given	0	0	0	0	62	**62**	*1.5*
Total	**1862**	**1673**	**273**	**218**	**122**	**4148**	-
%	*44.9*	*40.3*	*6.6*	*5.3*	*2.9*	-	-

permission for Dogslife to contact them by telephone and email. The majority (122/207) of these un-contactable people also declined to receive the newsletter but overall, 84.2% (3,626/4,307) owners opted to receive the newsletter.

2.3.3.5 Household type and other pets

Associations between different types of household and their pets are shown in Table 2.4. There were multiple associations between household type and the types of other pets reported. For example, families were disproportionately likely to have a cat and retired households disproportionately likely to have another dog.

2.3.4 Website Logins, Return Intervals & Loss To Follow-Up

Once registered, an owner could login to Dogslife website at their convenience and each login was recorded. Two owners had logged in 222 and 246 times respectively but the remaining owners had logged in fewer than 100 times. The distribution of number of logins, excluding the two outliers, is shown in Figure 2.4.

Logins to the website did not directly correspond with answers to the questionnaire (also referred to as 'making a data entry') because owners were able use the website for other purposes such as creating their scrapbook. Owners of 598 dogs registered with Dogslife but never answered the questionnaire and the

Table 2.3: Owner contact preferences according to household type

Contact Method	Permitted To Use	Family (%)	More than one adult (%)	Retired (%)	Single adult (%)	Not given (%)	**Total**
Email 4148	Yes	1742 (93.6)	1564 (93.5)	260 (95.2)	199 (91.3)	110 (90.2)	**3875** (93.4)
	No	120 (6.4)	109 (6.5)	13 (4.8)	19 (8.7)	12 (9.8)	**273** (6.6)
Phone 4064	Yes	564 (30.3)	559 (33.4)	125 (45.8[a])	84 (38.5)	80 (65.6[b])	**1348** (33.2)
	No	1298 (69.7)	1114 (66.6)	148 (54.2[a])	134 (61.5)	42 (34.4[b])	**2716** (66.8)

[a] $\chi^2 = 20.96$ (1df), $P < 0.001$
Red text indicates positive association.
[b] If an owner does not complete this question, permission is given by default.

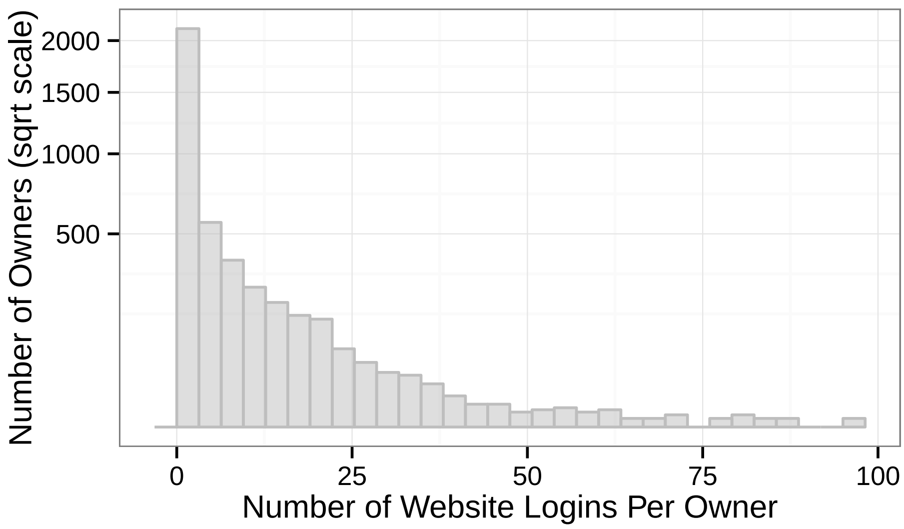


Figure 2.4: Owner logins to Dogslife website. Two owners who logged in more than 200 times each were excluded.

Table 2.4: The relationship between pet ownership and household type for Dogslife participants. Households that reported owning another dog, cat, other pet or did not report any pet (beyond their Dogslife registered dog), have been categorised by household type. Percentages are the percentage of each household type that reported having that type of pet. Individual households may appear up to three times in the table as they may, for example, own another dog, a cat and another pet.

	Another Dog *(%)*	Cat *(%)*	Other *(%)*	Dogslife dog only *(%)*
Family	521 (28.0*)	507 (27.2*)	430 (23.1*)	613 (32.9*)
More than one adult	564 *(33.7*)*	334 *(20.0)*	174 *(10.4*)*	767 *(45.8*)*
Retired	110 *(40.3*)*	41 *(15.0)*	9 *(3.3*)*	134 *(49.1)*
Single adult	84 *(38.5)*	36 *(16.5)*	24 *(11.0)*	92 *(49.1)*
Not given	5 *(4.1*)*	4 *(3.3*)*	4 *(3.3*)*	112 *(91.8*)*
Total	**1284** *(30.9)*	**922** *(22.2)*	**641** *(15.4)*	**1718** *(41.0)*

* χ^2 with Bonferroni correction indicates association, P <0.0025. For example, 28% (521 of 1862) of families reported having another dog compared with 33% (763 of 2286) for all other household types combined.
Blue text indicates negative association and red indicates positive association.
'Other' refers to all other types of animals listed in Section 2.3.3.2

owners of a further 460 dogs started but failed to complete a questionnaire.

Figure 2.5 shows a survival plot of time from registration for each dog. Dogs whose owners did not start or who failed to complete a questionnaire were considered censored and others were censored when they stopped keeping up to date with questionnaire answers. For dogs under one year this was considered to be after two months and for dogs over one year, this was four months from their most recent data entry. The ongoing recruitment means that very few dogs appear to survive until 42 months because the majority were too young. In reality, 1,161 dogs (26.9%, 95% CI: 25.6 - 28.3%) were up to date on 31st December 2013. This increased to 35.7% (95% CI: 34.1 - 37.4%) when the 1,058 dogs with no associated data entry were discounted. As the dogs reached one, two and three years of age, 44% (1432/3255), 35% (722/2093) and 29% (235/822) respectively were up to date. These values increased to 60% (1432/2474), 43% (722/1692) and 36% (235/652) when the group of 1,058 dogs were excluded.

The relatively conservative estimates of timing of loss to the project were belied by the true return intervals for the cohort which are shown in Figure 2.6. Some owners returned nearly three years after their last data entry. For younger dogs, 11% (95% CI: 4 - 18%) of return intervals fell beyond the two month interval and for the older age group, 23% (95% CI: 11 - 36%) fell beyond four months. The modal return intervals were 37 and 90 days respectively.

Time to assumed loss from the project was investigated using a Cox proportional hazards model and the results are shown in Table 2.5. The ability to contact owners by telephone or email improved the likelihood of their dogs being retained within the project. There were 78 owners associated with 84 dogs who refused all contact permissions but still came back and answered the questionnaire more than once per dog. Less obviously, people who reported that they were part of families or were smokers were more likely to be lost to the project. Oddly, newsletter subscription appeared to be associated with an increase in likelihood of loss to the project (hazard ratio = 1.3, 95% CI: 1.2 - 1.5). It was thought that this was perhaps because the default option for this field is to receive the newsletter. Therefore the owners who failed to complete registration, leaving the project very early, would be disproportionately likely to receive the newsletter. However, if these owners were removed from the analysis, the newsletter was still associated with loss to Dogslife. Owner location according to country was not associated with loss to Dogslife.

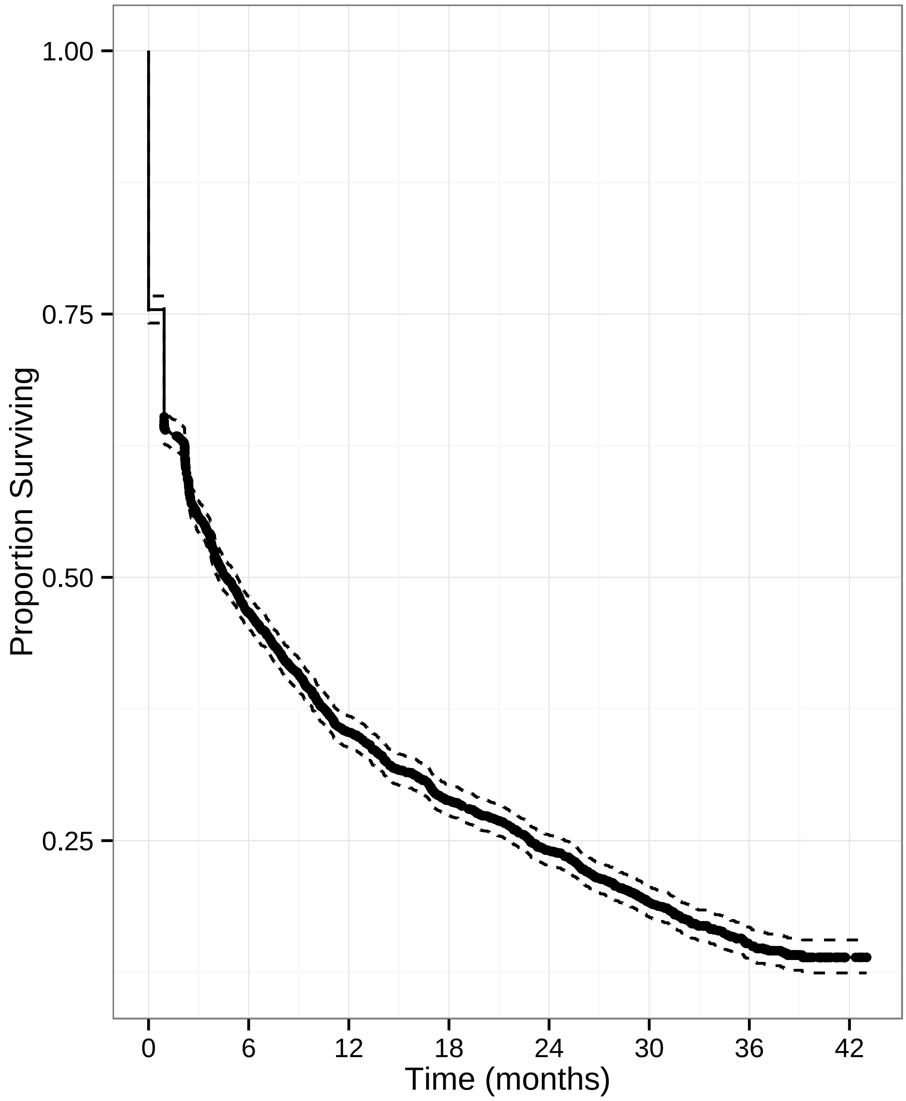


Figure 2.5: Time from registration to ‘loss’ from Dogslife with 95% confidence intervals

Table 2.5: Cox proportional hazards model of time to assumed loss to Dogslife

	Hazard Ratio e^{β}	95% CI: lower	upper	P-value
Household Types				
family	1	-	-	-
more than one adult	0.78	0.72	0.84	<**0.001**
retired	0.49	0.41	0.58	<**0.001**
single adult	0.83	0.70	0.99	**0.03**
not given	1.16	0.66	2.05	0.60
Smoking Status				
non-smokers	1	-	-	-
smokers	1.21	1.10	1.33	<**0.001**
not given	0.36	0.17	0.79	**0.01**
Postcode				
full postcode	1	-	-	-
first half only	0.68	0.23	2.00	0.49
not given	3.85	2.15	6.88	<**0.001**
Communications				
no telephone contact	1	-	-	-
telephone contact	0.56	0.51	0.60	<**0.001**
no email contact	1	-	-	-
email contact	0.44	0.38	0.50	<**0.001**
no newsletter subscription	1	-	-	-
newsletter subscription	1.30	1.17	1.45	<**0.001**
Other Household Pets				
no other dog	1	-	-	-
another dog	0.83	0.77	0.90	<**0.001**

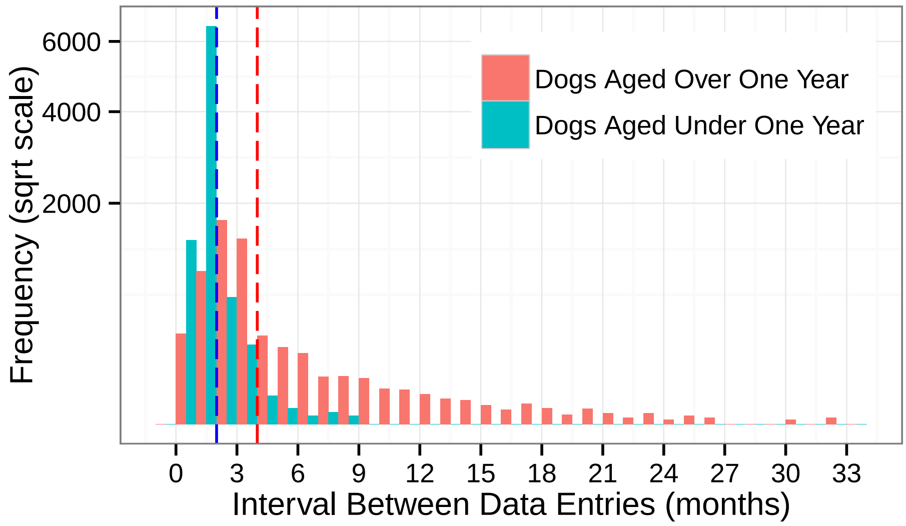


Figure 2.6: Intervals between data entries. Conservative cut-off times are shown using dashed lines (blue for dogs under one year and red for dogs over one year).

2.4 Discussion

Dogslife successfully recruited thousands of owners during the first three and a half years of the project. They were well described in terms of their household type, smoking status, location and whether they owned other pets. Some people did not give personal details about their households despite going on to give information about their dogs and one might speculate that these people either had strong feelings about personal privacy or suffered some sort of technical problem with the website during that stage of registration. Even the absence of information was useful and contributed to analyses of retention. Overall, the household information would be invaluable as a guide to environmental influences on the dogs throughout the project.

In order to generalise findings from the cohort, all data collected data must be considered in the context of potential selection bias. Participants were disproportionately likely to be female. Males are often under-represented in surveys, for example Søgaard et al. [2004], so this imbalance is not atypical of a study whose participants were self-selecting. It appears from viewing the postcode area recruitment map (Figure 2.3) that there were recruitment peaks in Shetland, the Western Islands and the City of London but each area had very low numbers of LRs registered with the UKKC meaning these rates had very large confidence intervals. Overall, Dogslife members were geographically distributed in proportion to LR UKKC registrations for whom address details were available and Dogslife household smoking rates were comparable to that reported for individuals in the UK. There was little evidence in terms of demographic factors that the recruited Dogslife cohort were unrepresentative of LR owners in the UK.

Retention bias was potentially more problematic as owners were being disproportionately lost to the project. People who described their households as families or whose household included a tobacco smoker were more likely to be lost to follow-up (Table 2.5). By contrast, retired households and those including another dog were more likely to be retained. Indeed, these two factors were themselves positively correlated within the cohort. In their examination of biases in a Spanish cohort study, Alonso et al. [2006] found a similarly increased risk of loss with regard to smokers and also that older people were more likely to be retained.

Piloting indicated that answering the Dogslife questionnaire took 5-10 minutes which is not a great burden but repeatedly answering the questionnaire and, in particular, measuring height and weight could be seen as onerous. As discussed in Clements et al. [2013], there was evidence that owners dropped out of the questionnaire when they were asked for their dogs' heights and weights. One might speculate that owners with limited spare time would be more likely to drop out and that owners in family households would have less time to give to Dogslife than those in retired households. The positive association between having another dog in the household and increased time in the project might be related to an increased interest in canine health or an acceptance that more of that owner's life would revolve around their dogs - including Dogslife participation.

Considerable efforts, such as the newsletter, prize draws, scrapbook facility and featured dog function, encouraged continued participation. Where contact was permitted, both automatic and non-automatic reminders were used to keep bringing owners back to the website. Dogslife employed many elements of best practice regarding increasing levels of questionnaire response (Edwards et al. [2009]) and thousands of owners reported information about their dogs more than once. Nevertheless, analyses based on data collected from the cohort would need to considered in the context of the validity of the data and the potential impact of retention bias. Data validity is addressed in the next chapter and the potential for retention bias is noted throughout this thesis.

Chapter 3

Data Validation

3.1 Introduction

Epidemiological data regarding exposures and outcomes that are not directly observable are typically collected by questionnaire. Re-using questionnaires that have already been validated minimises unnecessary effort on the part of investigators and future meta-analyses would be facilitated if standard questionnaires were available for different exposures or conditions. In human medicine considerable effort has been invested in designing and validating questionnaires that assess exposures including diet (Kaaks and Riboli [1997]), alcohol intake (Sieri et al. [2002]) and smoking (Leffondre et al. [2002]), and outcomes such as pain (Smith et al. [1997]) and depression (Radloff [1977]).

There are fewer examples in the veterinary field but attempts have been made with regard to specific species and syndromes. Hercock et al. [2009] attempted to develop and assess the validity of an owner-questionnaire addressing the extent of lameness in dogs with osteoarthritis of the elbow. Hotchkiss et al. [2006] followed the example of asthma questionnaires in human health to develop a risk-screening questionnaire for Recurrent Airway Obstruction (RAO) in horses. Investigators combined and weighted the horse-owners' answers to the questionnaire to create an overall risk profile for each horse and validated the questionnaire by asking participating vets to assess the horses' RAO status using more invasive tracheal wash or bronchoalveolar lavage techniques. The study involved 80 horses and data collection alone took three years indicating the effort required to develop such questionnaires. In veterinary medicine more broadly, different husbandry systems applied to multiple species mean that replicating the standard exposure questionnaires available in the human field is impractical. Instead specific questionnaires are developed for each new study and it falls to the designers to validate their questionnaire and assess the reliability and accuracy of resultant data.

For each new questionnaire it is necessary to determine whether the answers provide data which accurately reflect the exposure or outcome status of the subjects involved. Where possible, questionnaire answers should be compared with gold-standard test results which are administered concurrently. For more abstract concepts such as fitness, that do not have gold-standard tests, questionnaires often address proxies such as exercise levels. These questionnaires must then be assessed in terms of the extent to which reported exercise levels agree with other non gold standard assessment tools.

The Dogslife questionnaire was written specifically for the project by experts in canine health, epidemiology and genetics. Owners enter answers relevant to their dog(s) via an online platform (*www.dogslife.ac.uk*) that was piloted with approximately 40 dog owners, breed club officials and veterinary professionals prior to launch. The reported data reliability and accuracy are affected at two levels - firstly whether the owners understand the question as meant and secondly errors that they make in data entry. The need to understand and maximise the extent to which Dogslife data reflect the experience of the dogs as addressed by the online questionnaire underlies all future work and was the first issue addressed within the project. Aspects of what will be reported in this chapter have been published as part of a series of guidelines for implementing validation studies of internet-based longitudinal study data (Pugh et al. [2015c]).

3.2 Methods

Two different investigations were undertaken to determine the Dogslife data quality. Firstly, a random selection of Dogslife members were visited to measure their dogs, to check the accuracy of owner answers to the online questionnaire as seen in the Dogslife record and to garner owner opinions on the questionnaire ease of use. Secondly a set of veterinary records from another random sample of the cohort were collected. These records were checked against the Dogslife records of data given online by owners to determine whether the Dogslife illness and vaccination data were consistent with the veterinary records, gave a complete picture of the dogs' health and were sufficient to determine presenting signs and potential diagnoses described in the records. The questionnaire presented in Appendix 2 was amended as a result of the validation work so the original questions are detailed throughout this chapter.

3.2.1 Visits

One hundred dogs were selected at random from those whose owners had made a data entry in the six months prior to the date of selection (7th March 2012). The owner had to have provided a UK mainland postcode and not have formally withdrawn from the study. An article was then included in April 2012's Dogslife newsletter informing subscribers that one of the Dogslife team might be in touch to organise a visit of randomly selected dogs. The article was intended to facilitate recruitment by inciting interest and mitigating the negative impact of simply cold-contacting owners. Three of the one hundred owners selected did not give permission to be contacted by email and one had recently reported a family bereavement to the project team so was excluded. Of the 96 emails sent, one automatically bounced. Sixty-five owners responded by email or telephone. Three other owners (not in the 100 selected) also contacted Dogslife following the newsletter article and volunteered to be visited.

An initial itinerary was created based on early replies, and where feasible logistically, further visits were added to that itinerary as owners continued to get in touch. Over a six-week period from 23rd April 2012, 44 dogs were visited belonging to 43 different owners (including two of the three owners who volunteered and one randomly selected owner who had 2 dogs). Geographically, visits were undertaken across England, Wales and Scotland as shown in Figure 3.1.

A modified version of the online questionnaire (available as Appendix 3) was used for member visits and was intended to serve four purposes:

- Checking basic information, such as household type
- Allowing the visitor to view vaccination cards and measure the dogs
- Ascertaining the ease of use of the online questionnaire
- Facilitating discussion of the Dogslife study protocol

Figure 3.1: Locations of visited owners

When owners first register with the Dogslife website they are asked a series of one-off questions. It is possible for owners to change these answers by specifically clicking on "Edit your profile" on the Dogslife homepage but, beyond periodic articles in the Dogslife newsletter, they are not prompted to do so. On the basis that owners may not have updated their profiles, the visit questionnaire not only asked for their current situation but also whether that situation had changed. Their answers were then compared to those in the Dogslife database for 43 of the 44 dogs visited (excluding the second dog in the household of two Dogslife Labradors).

3.2.1.1 Dog Weights & Heights

Weights were measured using a Nintendo® Wii™ Balance Board (Nintendo Europe: Nintendo Center, PO Box 1501, D-63760, Germany) connected via blue tooth to an Apple® MacBook (Apple: 1 Infinite Loop, Cupertino, CA 95014, USA) using the WiiScale i386 application for Mac OS 10.5. The application did not produce a single weight measurement for any of the dogs due to their continued movement. Instead three screen-shots were taken of the application

Table 3.1: McBride's [2005] suggested descriptive scale for Lin's [1989] concordance measures (ρ_c)

Value of ρ_c	Descriptive Scale
<0.90	Poor
0.90 - 0.95	Moderate
0.95 - 0.99	Substantial
>0.99	Almost Perfect

whilst the dog was on the scale and these measurements averaged. This method of obtaining weight measurements was devised using dogs of various sizes prior to the visits and in testing it was determined to yield measurements within 0.1 kilograms (kg) of an Avery Weigh-Tronix pet scale (ITW Limited Foundry Lane, Smethwick, West Midlands B66 2LP).

The average visit weight for each dog was compared to the most recent weight entry in the Dogslife database given prior to the day of the visit. Concordance between the two measures was determined (Lin [1989]) and assessed according to McBride's [2005] suggested descriptive scale (Table 3.1). Owners are requested to input the weight of their dog at every data entry session regardless of the dog's age but it is the one element of the main online questionnaire that is voluntary due to the potential difficulties and possible health risks of weighing a large dog. There could be considerable delay between Dogslife and visit measurements. For comparison purposes, the dogs were divided by age according to whether they were over or under one year of age at time of visit on the premise that their weights would likely not change as rapidly after one year.

During the visits, the dogs' heights were measured using a seca213 stadiometer (seca UK: Medical Scales & Measuring Systems, 40 Barn Street, Birmingham, B5 5QB) comprising a vertical rule and an attached plate. The dogs were measured to the shoulder by standing them with their shoulders squarely above their front paws and lowering the sliding plate of the stadiometer to rest on the shoulders.

When Dogslife was created, owners were only asked to submit height measurements every month until the dog reached one year. The height measurements taken during visits were compared with the most recent measurement found for each dog in the Dogslife database that had been submitted on or before the day of the visit. Visit measurements of dogs over one year were expected to vary around the submitted measurement according to variation in measurement quality. For dogs under one year of age, given that the dogs were assumed to still be growing, it was assumed that the visit measurements should typically be equal or higher than those submitted prior to the visit. The comparison of submitted and directly measured data was therefore handled separately for dogs under and over one year.

3.2.2 Veterinary Records

An initial effort to collect veterinary records was undertaken from March - June 2012. One hundred dogs were randomly selected whose owners had indicated that Dogslife could contact them by email or telephone. Those who had asked to leave the study were excluded and given that approximately 25% of owners register their dog with the project but never complete a data entry, only those who had made at least one data entry were eligible for selection. Where permitted a single attempt was made to contact the chosen owners by telephone to ask whether Dogslife could request their dog's veterinary record. If telephone contact failed or telephone contact was not permitted, a single email was sent. If the owner agreed, a consent form was sent to them and if they signed and returned it, the vet(s) were contacted once to request the dog's full clinical history. To increase the number of records obtained, all visited owners were also asked to complete the form. Two owners were on both the list to be visited and the list of those to contact regarding veterinary records and had already completed their forms before the visits took place. All telephone contacts were undertaken by the project administrator who has a strong background in phone-based communications and has developed a rapport with many members since Dogslife's inception.

The veterinary records contained information regarding vaccinations, neuter status, weight, illness incidents and partial information regarding worming and flea treatment products. Initial comparisons of prophylactic worming and flea treatments, neuter status and dog weight were undertaken using 17 records. Vaccinations and illness incidents were examined more quantitatively using all 66 records that had been collected in the period March - June 2012. With only 66 veterinary records the frequencies of different types of vaccination and illness were too low for useful analysis. On the basis that it would be desirable to understand and quantify the extent of illness under-reporting, further veterinary records were sought.

Table 3.2: Stage by stage success of obtaining vet records

Group (comprising)	Contact Method	Number Agreed	Signed Forms Returned	Records Obtained	Success Rate
1st	73 emailed	30	23(1)	23(1)	32%
100(3)	20 phoned	20	14(2)	14(2)	70%
	7 uncontactable				
2nd	83 emailed	31	26(3)	26(3)	36%
142(9)	45 phoned	43	31	31	69%
	14 uncontactable				
Visited*					
43(1)		41	41(1)	38(1)	93%
Total					
285(13)		**165**	**135(7)**	**132(7)**	**46%**

Where the owner was uncontactable, this implies phone contact was attempted but failed. The numbers given pertain to households. Numbers in brackets refer to second and third dogs in Dogslife households.
* Two of the visited owners had already returned forms and are included in the 1st group.

In November 2012, in addition to re-starting the effort to contact the initial tranche of owners, the owners of a new group of 142 dogs were randomly selected. The same criteria were used with the exception that the owners had to have made at least two data entries rather than the one required of the initial group. Two data entries would increase the period of time during which any illnesses would be expected to be reported to Dogslife - this change prevented the further collection of records pertaining to dogs whose owners had only made one data entry, effectively having a very limited 'at risk' period. Where the owners of chosen dogs had other Dogslife dogs in their household, records were also sought for these dogs on the basis of convenience. The numbers of owners agreeing at each stage of the process are shown in Table 3.2.

3.3 Results: Visits

The 100 dogs initially selected for visits were distributed as shown in Table 3.3 (for completeness, the precise options offered to owners in the questionnaire are given below in Section 3.3.1).

Table 3.3: Comparison of the composition of the cohort at time of selection and the selected and visited samples

		Cohort n=2,255	Selected n=100	Visited n=43
Dog Sex	Female	48%	54%	61%
	Male	52%	46%	39%
Coat Colour	Black	49%	46%	56%
	Yellow	26%	32%	22%
	Chocolate	22%	20%	20%
	Other	2%	2%	2%
Household Type	Family	48%	46%	41%
	More Than One Adult	42%	37%	32%
	Single	6%	8%	12%
	Retired	7%	8%	15%
	Not Given	4%	1%	0%

The distribution of dog ages on the days of their visits are shown in Figure 3.2. The dogs ranged in age between 150 and 825 days with 15 dogs ages under one year. The mean age was 476 days (95% CI: 415 - 537 days).

3.3.1 Household Profile

3.3.1.1 Household Type

When registering owners were asked "How would you describe your household?" and offered five original options: Family (one of more adults and one or more children); More than one adult; Single adult; Retired (single or couple); Other.

None of those visited had chosen 'Other' but two of the answers given during the visits were ambiguous - both with relation to children. One household included adult children and could be considered as 'More than one adult' but the owner described them as a 'Family' and another household included children at weekends and could be placed in either category. Of the remaining 41 owners, 37 visit and database answers agreed. Two owners had answered 'Single adult' when registering but when visited they mentioned that they were 'Retired' and two others had described their households as 'Families' in the database but when speaking with them when visiting it became apparent that 'More than one adult' would be more appropriate.

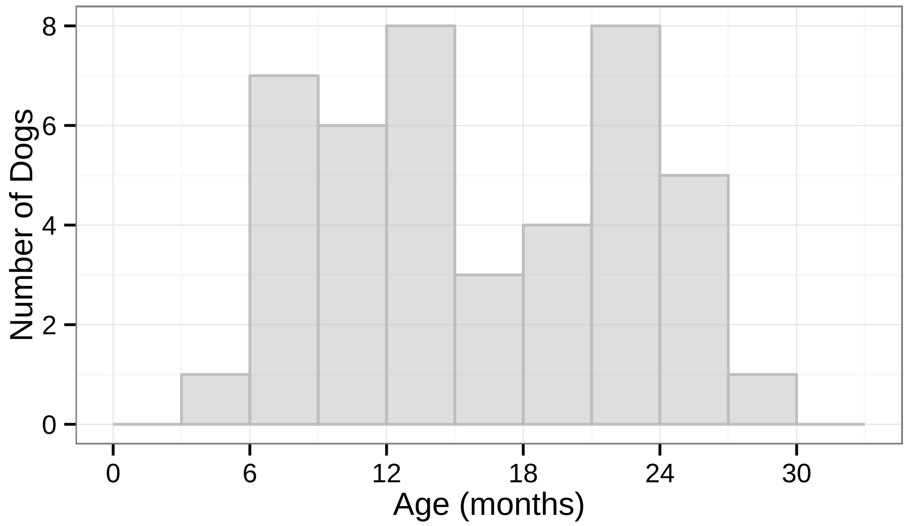


Figure 3.2: Dog age when visited

Table 3.4: Comparison of household smoking status at registration and visit

		Visit Response	
		Smoking	Non-smoking
Dogslife Record	Smoking	3	1
	Non-smoking	4	33

3.3.1.2 Smoking

Of the 43 owners visited, one was not asked the smoking question (available in Appendix 1) during the visit and another mentioned that they had recently given up. The owner who had given up had originally answered 'Yes' when registering and had not changed the answer since quitting. The answers of the remaining 41 owners are shown in Table 3.4. The overall Kappa score for agreement is a moderate 0.48 (standard error (se) = 0.19) but it appears that there is a significant degree of misclassification amongst smokers. During the visits several people answered 'Yes, but not in the house' and this may partially explain why the four owners who had answered 'No' to the smoking question online but 'Yes' during the visit.

Table 3.5: Comparison of other pets in the household at registration and visit

		Visit Response			
		Dogs	Cats	Other	None
Dogslife Record	Dogs	16	0	0	0
	Cats	0	11	0	0
	Other	0	0	2	2
	None	1	0	1	13

3.3.1.3 Other Pets

As discussed in Section 2.3.3.2, owners are asked about other pets in the household at the end of registering (question available in Appendix 1). Table 3.5 shows the types of other pets that lived with visited dogs. Five owners mentioned that the number or type of other pets had changed since joining the project and of these, one had updated her profile. The other four were correct at the time of registration but incorrect at the time of the visit. Two owners did not mention any changes but their visit answers were different to those originally entered into the database.

3.3.1.4 Dog Purpose

When first answering the questionnaire, as a one-off question, owners were asked about their dog's primary purpose (question available in Appendix 2, Section A2.2). The forty three dogs included in this analysis comprised forty household pets, two assistance dogs and one working dog; there was complete agreement between Dogslife and visit answers.

3.3.2 Repeated Dogslife Questionnaire

The household profile was captured by the registration process and during the first questionnaire entry but descriptions of the dogs' lives were captured by repeated answering of the Dogslife questionnaire. Where seasons and breeding are concerned, the questionnaire is animal sex specific. Where the word Dog* is used, the dog's name is automatically inserted.

Table 3.6: Comparison of sleeping locations reported online and in person

		Visit Response			
		Alone	Shared (family)	Other (shared dog)	Other
Dogslife Record	Alone	22	0	3	0
	Shared (family)	0	7	0	0
	Other (shared dog)	0	0	7	0
	Other	0	0	1	0

3.3.2.1 Sleeping Location

Owners are asked "Where does Dog* sleep at night?" and the original answer options were as follows: Alone in a room in a house; In a room shared with a member of the family; Outside; Other. None of the dogs visited slept outside but several of the owners mentioned their dissatisfaction with the possible answers. As shown in Table 3.6, eight owners had entered 'Other' rather than 'Alone' online because their dog slept with another dog. Of those, seven specified this in the available free text box and the eighth simply left it blank. Three owners entered 'Alone' into the database but when visited mentioned that their Dogslife dog slept with another dog. In light of the difficulty owners were experiencing answering this question, the potential answers were amended on 3rd October 2012 to the those now shown in Appendix 2 (Section A2.3).

There were also three more ambiguous answers given during the visits (not included in Table 3.6). Two had answered 'Alone' in the database but both owners mentioned that the dog would occasionally sleep with them and the third said that their dog had slept outside before coming to them, then with their other dog and then alone. The Dogslife record reflected the two locations during their ownership i.e. with another dog and then alone.

3.3.2.2 Diet

Owners are asked several questions with relation to diet (questions available in Appendix 2, Section A2.5). The original options for how often the dog was fed were: Once daily; Twice daily; Three times daily; More than three times daily. The potential answers regarding timings were originally: In the morning; In the evening; In the morning and evening; Throughout the day; Multiple times throughout the day. The final question relating to titbits originally asked "Does

Table 3.7: Feeding Frequency

		Visit Response (daily)			
		Once	Twice	Three times	More than three times
Dogslife Record	Once	2	0	0	0
	Twice	0	36	1	0
	Three times	0	0	3	0
	More than three times	0	0	0	1

Dog* also receive titbits or left-over food from your own meals and snacks?" [Yes/No].

The feeding frequencies are shown in Table 3.7 but again there was some frustration with the available options. One of the dogs that was fed once daily was fed in the evening but the other was fed *ad lib*. The owner of the latter dog chose 'Throughout the day' in the timings question but felt that this should also be available as a frequency option. In light of her comments, the possible answers to feeding frequency were amended to include 'Throughout the day'.

All thirty-six owners who answered 'Twice' fed their dogs in the morning and evening. One dog whose Dogslife record indicated that they were fed 'Twice' daily had had their regime changed to 'Three times' daily but the owner had not yet updated the Dogslife record. In each case where the dog was fed 'Three times' daily the owners expressed some annoyance with the possible options for timings of feedings. They each entered 'Multiple times throughout the day' but wanted the option to put 'Morning, lunchtime and evening'. The option 'In the morning, lunchtime and evening' was subsequently added to the list of answers.

The final dog suffered from severe dietary intolerances and shortly before the visit the owners had finally found a diet and feeding regime he could manage. The dog's difficulties were documented in the illness sections of the database and his feeding regime at the time comprised four meals at six-hourly intervals. Within the database his feeding regime was captured as being fed 'More than three times daily' at the times 'Multiple times throughout the day'.

The distribution of food types is shown in Table 3.8 and where the diet was relatively simple such as dried or a mixture of dried and tinned food the answers had good agreement. Owners appeared to have more difficulty choosing a category when the diet was more varied. For example, there were four owners

Table 3.8: Comparison of food types reported online and in person

		Visit Response				
		Dried	Tinned	Mixture	Home prepared	Other
Dogslife Record	Dried	27	0	0	1	3
	Tinned	0	0	0	0	0
	Mixture	0	0	7	0	0
	Home prepared	0	0	0	1	0
	Other	0	0	0	0	4

who had described their dog's diets as 'Dried' online but elaborated in person to describe a diet of dried food plus meat, vegetables, rice, gravy, and fruit. Unfortunately none of these additions were captured in the database. Three owners did mention online and in person that they added home prepared or other food into their dog's diet by ticking the 'Home prepared' or 'Other' option. In both cases a free text box appears that allows the owner to detail what they feed their dog. These three dogs were fed dried food plus either raw meat or tripe. The remaining two dogs (both owners answering 'Other' online and in person) were the dog with dietary intolerances who was fed Nature's Harvest wet food from packets and a dog who was fed raw food and dietary supplements.

When dried, tinned or a mixture of both are chosen, owners are asked to select the brand of food that they feed their dog from a drop-down list. Owners typically choose 'Other' from that list and then type the specific brand into the free text box provided. Thirty-one of the owners gave the same brand of dried food during the visit as was already in the database and five also gave the same brand of tinned or wet food. Fourteen owners gave different brands of dried or tinned foods. The most popular brand amongst the visited group was James Wellbeloved which was fed to six or seven dogs according to the database and visit answers respectively.

Figure 3.3 compares the weight of food given according to Dogslife and visit answers with linear regression lines for each food type. In total, only fifteen visit answers agreed with those already entered into the database with a further eight answers within 10% of the quantity in the database. Three answers (70 grams (g) and 85g; 125g and 150g; 300g and 340g) had relatively low absolute differences that could perhaps have been measurement differences. Two answers were precisely double that already in the database and two further were almost

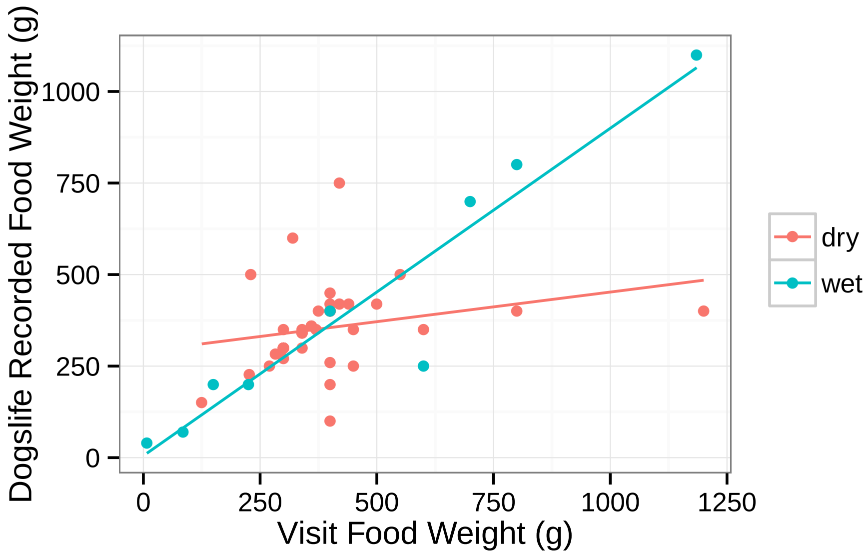


Figure 3.3: Weight of food reported to Dogslife and at visit. The blue and red lines are the linear regression lines for wet and dry food respectively

double; in all four cases the dog was fed twice per day indicating that perhaps those owners had entered the quantity given per meal rather than per day. In the remaining cases, the quantities given during the visits were markedly different than those given online. The discrepancy between answers is less evident with regard to wet food where the owner can simply read the weight from the side of a tin or packet.

In some cases the delay between data entries and subsequent visits may have included a change in feeding regime and indeed, the change between food weight at last data entry and food weight at visit is largely positive (Figure 3.4) indicating that food consumption had increased. However Figure 3.4 does not show a correlation between the weight difference and the delay between measurements. As mentioned previously, the online questionnaire is limited with regard to fully capturing the dogs' diets and this maybe reflected in the lack of agreement between visit and online answers. For example during the visit, one owner described giving 400g of wet food with 100g of dried food for two days each week and just 400g of dried food on all other days. This was particularly poorly captured in the database which simply said 400g of each every day.

The final food question regarding titbits was undoubtedly the most problematic. Twenty-two of the owners had previously answered 'No' to the titbit

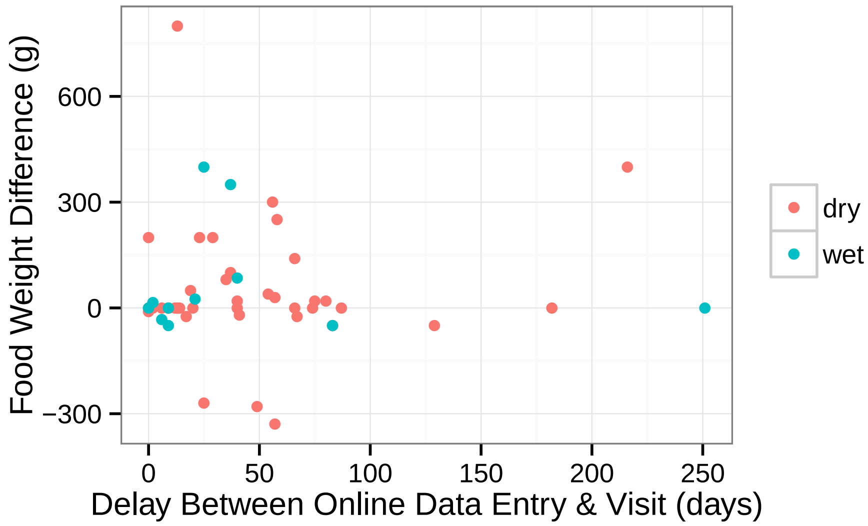


Figure 3.4: Delay between data entry and visit against change in food quantities reported

question online but during visit questioning 20 of these owners indicated that their dog did receive other food and the remaining two either were not asked or did not answer. The type of extra food was not explicitly asked in the visit questionnaire but in order to elicit an accurate answer to the yes/no question, a discussion regarding the nature of titbits was typically initiated. Among the 20 owners who had initially answered 'No' online but then said 'Yes' in person, the types of titbits they apparently fed their dogs was noted for 16. None of these 16 included food from human plates but the extra food included everything from their normal food as a training tool to apples, carrots, cheese and chicken. The remaining twenty owners all said online that their dog received titbits and in person mentioned left-overs of roast dinners, rice pudding, licking plates and simple dog chews.

3.3.2.3 Bathing

Table 3.9 compares the visit answers with the Dogslife database answers for bathing frequency (questions available in Appendix 2, Section A2.3). Many owners found it difficult to answer using the Dogslife categories and instead preferred 'Once', 'Twice', or 'When (s)he needs it'. Each dog can appear in the table more than once because owners gave multiple answers both during visits and online. In the cases of 'Once' and 'Twice', provided the owner had made a Dogslife answer of monthly or less than monthly once (or in the case of Twice, twice) then the answers were considered to agree. Four owners could not be more specific than 'When (s)he needs it' despite prompting. Dogslife and visit answers for the material or product used to bathe the dogs are shown Table 3.10. Several owners mentioned using the hose so this category was added to the results table (but not the questionnaire).

Table 3.9: Bathing Frequency

		Visit Response						
		When (s)he needs it	Once	Twice	Fortnightly	Monthly	Less than monthly	Never
Dogslife Record	Weekly	0	0	0	0	0	0	1*
	Fortnightly	0	0	0	1	1	0	1*
	Monthly	1	0	0	3	4	1	0
	Less than monthly	4	0	0	1	1	0	1*
	Never	0	0	0	0	0	0	13
	Agree	0	4	1	0	0	0	0

*These three entries refer to two dogs. Both owners made bathing entries online but mentioned during the visit that they did not consider their dog as having been 'bathed' because they just used a hose.
In the cases of visit responses of "Once" and "Twice", provided the owner had made an online answer of monthly or less than monthly once (or in the case of "Twice", twice) then the answers were considered to agree.

Table 3.10: Bathing Material

		Visit Answer					
		Water alone	Dog shampoo	Human shampoo	Other	Hose	Not mentioned during visit
Dogslife Record	Water alone	1	0	0	0	0	1
	Dog shampoo	2	24	0	0	3	0
	Human shampoo	0	0	1	0	0	1
	Other	0	0	0	1*	0	0
	Other (hose)	0	0	0	0	1	0
	Not mentioned in Dogslife	1	3	0	0	1	0

*During the visit the owner mentioned using medicated shampoo. In Dogslife they had chosen 'other' and detailed medicated shampoo in the free-text box.

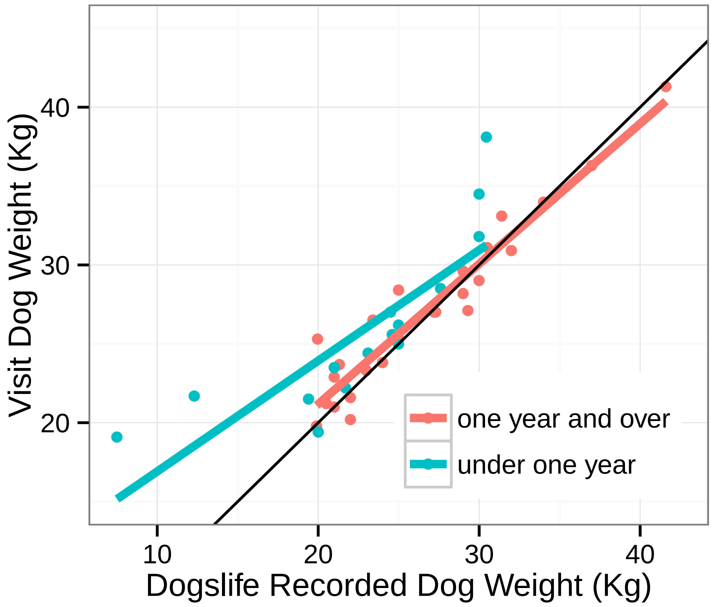


Figure 3.5: Comparison of visit and previously reported Dogslife weights (y = x line in black)

3.3.3 Dog Measurement: Weight

In total, 41 of the 44 dogs were weighed during the visits. Of the remaining three, one dog could not be caught for weighing, the Wii balance board did not work on one occasion, and one dog appeared to suffer pain when touched near its tail and it could not be induced to sit on the scale without such handling.

The average measured weight of the dogs aged under one year was 25.9kg (95% CI: 24.2 - 27.6kg, range = 19.1 - 38.1kg). The average magnitude difference between these measurements and the previously entered weights was 3.2kg (95% CI: 2.1 - 4.3kg, range = -0.6 - 11.6kg). Figure 3.5 shows the database measures against those measured during the visits in blue. The linear regression is markedly different from the simple x = y line shown in black. A single measurement was lower than those previously submitted and that visit took place only 6 days after the earlier submission. On average, the time between the prior weight being submitted and the visits was 31 days (95% CI: 23 - 38 days, range = 0 - 70 days).

The average measured weight of the dogs aged one year and over was 27.1kg (95% CI: 25.5 - 28.7kg, range = 19.8 - 41.3kg). These measurements were closer to the previously submitted values than for the younger dogs with an average

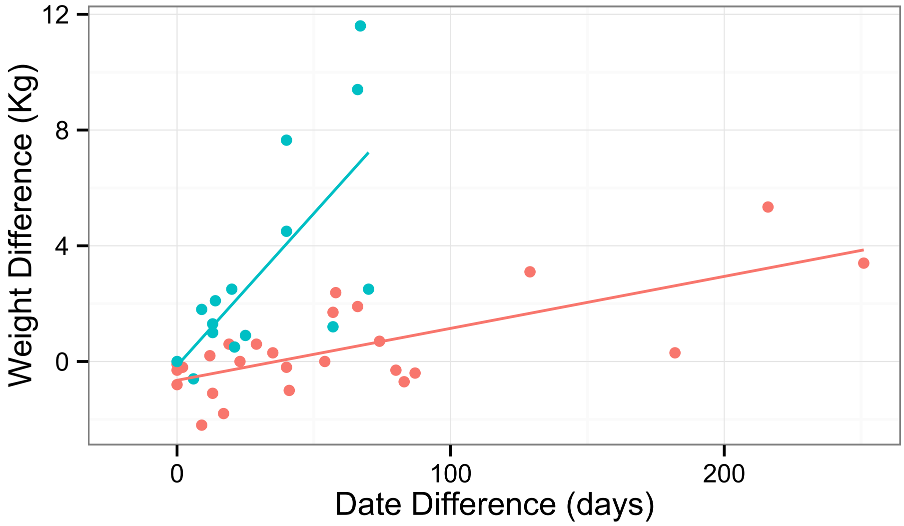


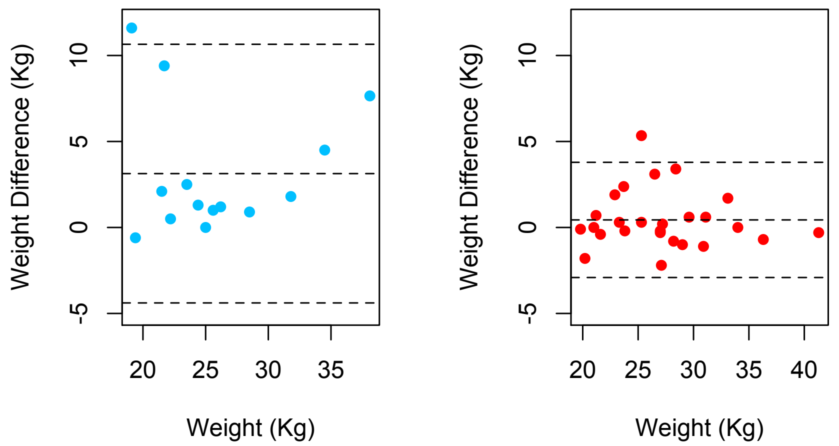
Figure 3.6: Weight measurement difference against delay between database entry and visit. Linear regression lines are included for dogs under one year in blue and dogs of one year and over in red.

magnitude difference of just 1.1kg (95% CI: 0.74 - 1.5kg, range -2.2 - 5.34kg). The red points in Figure 3.5 show the database measures against those measured during the visits. For these older dogs, the linear regression line is remarkably close to the simple x = y line shown in black. Lin's concordance coefficient for the two measures indicates agreement on the cusp between moderate and substantial at 0.95 (95% CI: 0.89 - 0.98) (Lin [1989]; McBride [2005]) (Table 3.1).

3.3.3.1 Summary: Weights

The average delay between prior weight submissions and weight measurements during a visit was 61 days (95% CI: 41 - 81 days, range = 0 - 251 days) and there is a trend between increased delay and increased difference in measurement (Figure 3.6). The younger dogs have been plotted in blue with a linear regression line that is considerably steeper than that of the older dogs plotted in red.

Figure 3.7 shows Bland-Altman plots for the younger and older dogs comparing weights measured during the visits with weights previously submitted by owners. The horizontal lines represent the mean difference ±2standard deviation (sd). The smaller sd in the older group is illustrative of better agreement between measures in the dogs over one year when the dogs have largely finished growing



[a] Dogs under a year [b] Dogs of a year & older

Figure 3.7: Bland-Altman Plots: Weight

rapidly and the delay between submitting measurements and subsequent visits is less significant.

3.3.4 Dog Measurements: Height

In prior testing, two people could reliably use the stadiometer to take dog height measurements in a veterinary setting but on visits many of the dogs were reluctant to stand to their full height and appeared nervous of having something unfamiliar behind their heads. Nevertheless heights were measured for all 44 visited dogs.

One of the 15 dogs under one year did not have any reasonable height measurements in the database. There were only two measurements entered and both were too low to be accurate but too high to be the subject of a simple metric vs. imperial error. This dog was therefore excluded from the analyses of height. The average measured height was 57.7 centimeters (cm) (95% CI: 56.0 - 59.3cm, range = 50.5 - 62cm). The blue points on Figure 3.8 show the database measures against those measured during the visits for dogs under one year and the linear regression line is markedly different from the simple x = y line shown in black. As expected all measured heights were higher than those entered into the database with an average increase of 5.0cm (95% CI: 3.4 - 6.6cm, range = 1.0 -

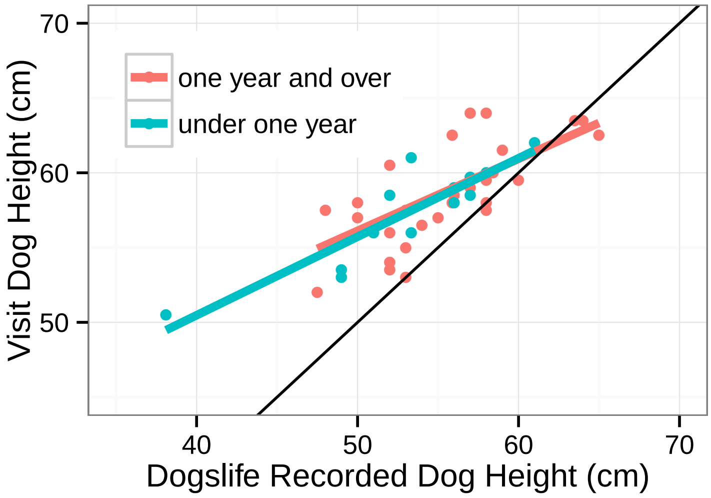


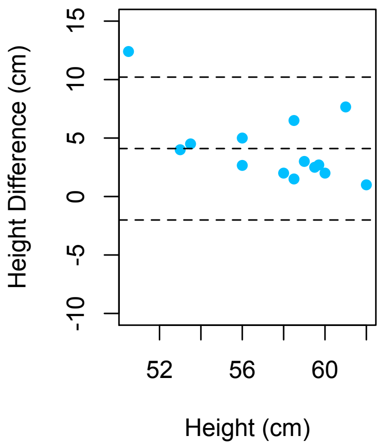
Figure 3.8: Comparison of visit and previously reported Dogslife heights (y = x line in black)

12.4cm). The elapsed time between owner entries and visits was an average of 28 days (95% CI: 17 - 39 days, range = 0 - 67 days). The gap of 67 days was associated with the greatest change in height of 12.4cm indicating the desirability of frequent data entry by owners to capture rapid changes.

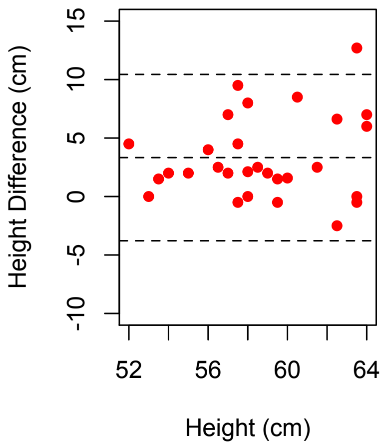
The average measured height for dogs of one year and over was 58.7cm (95% CI: 57.5 - 60.0cm, range = 52 - 64cm). The average magnitude of difference between measurements taken during a visit and those given by owners was 3.6cm (95% CI: 2.4 - 4.8cm, range = -2.5 - 12.7cm). The two greatest differences 9.5 and 12.7cm both related to dogs whose owners had not entered height measurements when the dogs were one year of age. Instead the ages at last height entry were 176 and 225 days respectively. Figure 3.8 shows the database measures against those measured during the visits in red and, again, the linear regression line is markedly different from the simple x = y line shown in black. Lin's concordance coefficient for the two measures indicates minimal agreement at 0.45 (95% CI: 0.21 - 0.64) (Lin [1989]; McBride [2005])(Table 3.1).

3.3.4.1 Summary: Heights

The visit measures and difference between these measures and those entered by owners are shown below in Bland-Altman plots for the younger (Figure 3.9 [a]) and older (Figure 3.9 [b]) dogs. As before the horizontal lines represent the mean



[a] Dogs under a year



[b] Dogs of a year & older

Figure 3.9: Bland-Altman Plots: Height

difference ±2sd but age group is not the the only guide to difference between visit and database measurements. Figure 3.10 demonstrates that delay between the timing of the database measurement and subsequent visit also had a strong effect on the difference between measurements. The older dogs are plotted in red and the younger in blue and linear regression lines for each group are markedly different.

3.3.4.2 Illness

The illness section of the visit questionnaire was more a test of owner recall than data validation but it did highlight certain limitations with the illness page on the website (questions available in Appendix 2, Section A2.7). For example one dog consumed four packets of throat lozenges and required a vet visit and the prescription of emetics but this was not mentioned via the online questionnaire. The owner stated that they had not entered the information because of the title of that website page is "Illness" and they did not consider their dog to have been ill.

Several owners mentioned that they knew what had caused the clinical sign in their dog and wished they had the option to explain the issue online. For example

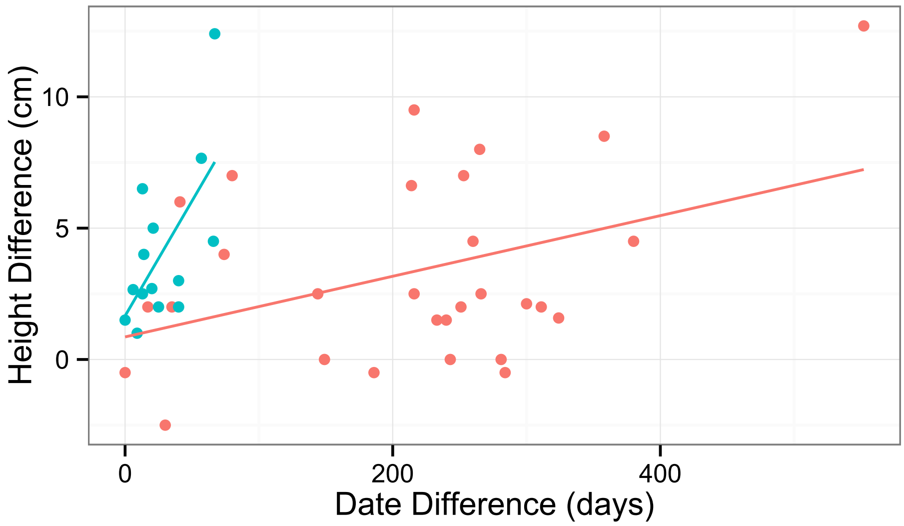


Figure 3.10: Difference in height measurements vs. delay between online entry and visit. Linear regression lines are included for dogs under one year in blue and dogs of one year and over in red.

one dog vomited each time he drank cows' milk and the owner had never entered these incidents because again they did not regard their dog as being ill; they simply stopped feeding him milk. In response to this feedback, and in the hopes of encouraging owners to tell Dogslife about incidents that the owners did not regard as illnesses, an extra free-text box was added to the questionnaire. If an owner indicated that their dog had been ill but that they had not taken it to the vet, a box entitled "Do you know why Dog* developed this illness?" was added on 1st February 2013.

In total, twelve owners showed complete agreement between their online Dogslife record and their visit questionnaire answers. Of the others, twenty mentioned something online that they forgot during the visit and twenty-three mentioned something during the visit that they had not previously entered into the website. For both types of non-agreement, gastrointestinal illnesses were heavily represented.

Dogslife & Vet Records	Number of Dogs	
	Worming	Flea treatment
Agree	7	2
Disagree	2	1
Incomplete information	5	8
No information	3	6

Table 3.11: Comparison of Dogslife and Vet Record worm and flea treatment brands for each dog

3.4 Results: Veterinary Records

Of the 139 records collected, five were disregarded. Two referred to dogs with only one Dogslife data entry and therefore minimal at-risk period, one was illegible (a poor facsimile) and two referred to mixed-breed dogs with incorrect dates of birth. It seemed likely that these latter two records did not refer to Dogslife dogs. Seven dogs had no illnesses in their veterinary and Dogslife records so these were disregarded for illness comparison. Six dogs had no vaccinations in their Veterinary records during the period of time when they were members of Dogslife so they were disregarded for vaccination comparisons. Initial comparisons were undertaken with 17 records to assess, for each element of interest, whether the records could sensibly used. More detailed assessments were undertaken with all 134 useable records.

3.4.1 Prophylactic Worm and Flea Treatment

Using veterinary records to determine timing of application and products used for worming and flea treatment is problematic. Owners may purchase the products in batches so whilst the brand of product might be identifiable, the date of administration is typically unclear. Comparison is further complicated by the fact that owners can buy medications elsewhere so the absence of a product from a veterinary record does not necessarily mean that the owner is incorrect in citing its use. With these caveats, Table 3.11 shows the degree of agreement regarding brand between Dogslife and the vet records of 17 dogs. The 'incomplete information' and 'no information' categorisation may apply to the Dogslife record, the vet record or both.

3.4.2 Neuter Status and Dog Weight

All 17 Dogslife records agreed with the vet recorded neuter status of the dogs with only one disagreement regarding the date of the neutering operation. This entry was incorrect by 31 days indicating a simple out-by-month error. Fifteen out of 17 records included one or more weight measurement(s) and all were consistent with the entries in the database.

3.4.3 Vaccinations

All 128 useable veterinary records from the first and second tranches of records collected were considered (exclusion criteria are described in Section 3.4). According to these records 370 vaccinations were given, of which 148 were given before the owner registered their dog for Dogslife and 20 were given after their most recent data entry. The latter 20 were disregarded as not part of a Dogslife 'at-risk' period but 92 of the 148 veterinary recorded vaccinations given before Dogslife registration were recorded in the Dogslife record (62%) so dismissing everything from this early risk period would have meant ignoring valuable information. Of the remaining 202 veterinary recorded vaccinations given between Dogslife registration and most recent data entry, 160 had been recorded in Dogslife (79%).

The online questionnaire asked owners whether their dog had been vaccinated since they last visited the site (or in the first instance, 'in the last four weeks') (questions available in Appendix 2, Section A2.6). The owner was only able to give one vaccination date each time they answer the questionnaire so if the dog had been vaccinated more than once since the last data entry, the owner could only report one of the instances. The typical interval between first vaccination and booster for puppies is just two weeks. Both vaccinations would only be reported if the owner made a data entry between vaccination and booster.

Examination of the veterinary records indicated that the majority of dogs were given initial routine vaccination courses comprising vaccination and booster at approximately 2 - 3 months of age (three dogs were also given a second booster). The vaccinations in the veterinary and/or Dogslife records are shown in Table 3.12. If more than one dose of a single type of vaccination was given between Dogslife data entries, the owner would only be able to give the dates of one instance (for example, they might be able to report a booster but not the original vaccination and *vice versa*). The various components of each vaccination have been

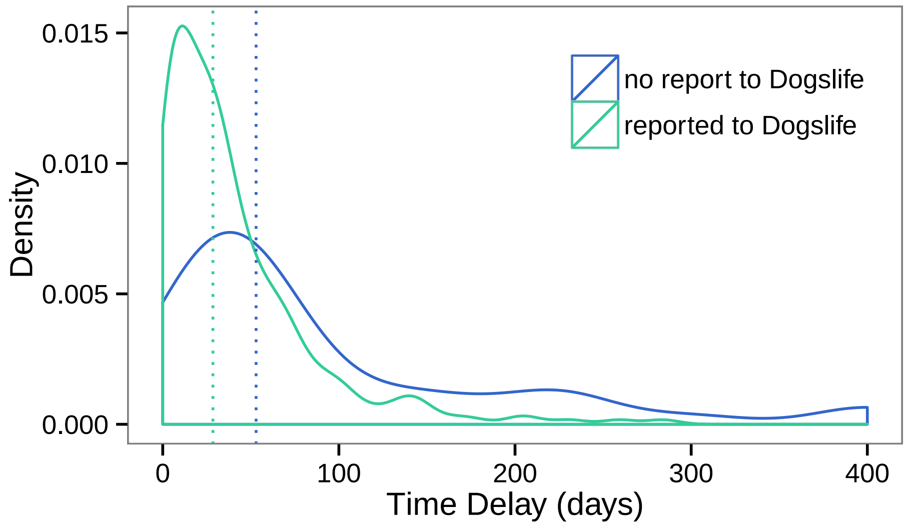


Figure 3.11: Time delay between vaccination and subsequent visit to Dogslife website. Dotted lines indicate median delay.

collapsed in Table 3.13 for easier interpretation of *truly* missing or added vaccinations. Table 3.13 included 76 multi-stage vaccinations comprising 155 individual vaccinations of which only six vaccinations were entirely unreported by owners.

The delay between vaccination and subsequent Dogslife data entry appears to be associated with whether an owner reports a vaccination. Figure 3.11 illustrates this relationship. For the vaccinations found in the veterinary records, the median delay between a veterinary visit and when it was recorded with Dogslife was 28 days (Inter-quartile range (IQR) = 42 days). By contrast, single vaccinations or courses of vaccinations that were omitted from the Dogslife record occurred a median of 48 days (IQR = 82 days) before the owner made a subsequent data entry.

Figure 3.12 shows the proportion of each type of vaccine according to the level of agreement between Dogslife and the vet records. There are too few rabies vaccinations to draw useful conclusions (data not shown) but comparisons between routine and kennel cough vaccinations are possible. There is no difference in the proportion of routine and kennel cough vaccinations seen in the vet record but missing from Dogslife. However, the Dogslife record is more likely to include kennel cough vaccinations that are not found in the vet record than routine vaccinations.

Table 3.12: Recording Of Vaccination By Timing & Type

		Vaccine							
Vaccine Timing	Recorded In	KC	Routine (single)	Routine (1st)	Routine (booster1)	Routine (booster2)	Rabies (single)	Rabies (1st)	Rabies (2nd))
Before Dogslife Registration	Vet Record	1	4	45	6	0	0	0	0
	Dogslife Record	4[a]	1[b]	0	0	0	0	0	0
	Both Records	7	28	20	35	2	0	0	0
Between Registration & Latest Data Entry	Vet Record	9	15	6	10	0	0	1	1
	Dogslife Record	9[a]	5[b]	0	0	0	2[c]	0	0
	Both Records	37	88	3	23	1	6	1	1

KC: Kennel Cough
Routine: Initial routine courses of vaccination comprise 'Routine (1st)', 'Routine (booster1)' and on three occasions 'Routine (booster2)'. All future routine vaccinations were given singly 'Routine (single)'.
Rabies: Rabies vaccination involves one injection followed by a titre test for the presence of sufficient antibodies to convey immunity. Where only one vaccination was given, 'Rabies (single)' is used. Where re-vaccination was required following a failed antibody test, 'Rabies (1st)' and 'Rabies (2nd)' were used.
[a] All 13 additional KC vaccinations found in Dogslife that were not in the veterinary records coincided with 'routine' vaccinations, given on the same day, that *were* found in the veterinary records.
[b] Two of the six additional 'routine' vaccinations found in Dogslife that were not in the veterinary records coincided with KC vaccinations, given on the same day, that *were* found in the veterinary records. Three of the six were simply repeat mentions of vaccinations already in the Dogslife and veterinary records that had slightly different dates. The final one of six was not found in the vet record but was recorded on the dog's vaccination record which was observed during a visit.
[c] One of the additional rabies vaccinations found in Dogslife that was not in the veterinary records coincided with a 'routine' vaccination, given on the same day, that *were* found in the veterinary records - the dog had previously been vaccinated for rabies. The other additional rabies vaccination was a repeated mention of a vacccination found in both records with a slightly different date.

Table 3.13: Number of vaccinations that agree with the vet record or have been missed or added by owners

	Agree	Owner Missed	Owner Added
Routine[1]	184	25	6
Kennel Cough	44	10	13
Rabies[2]	8	0	2

[1] Routine vaccinations comprised the first routine vaccination and all boosters.
[2] Rabies vaccinations comprised all attempts at immunisation.

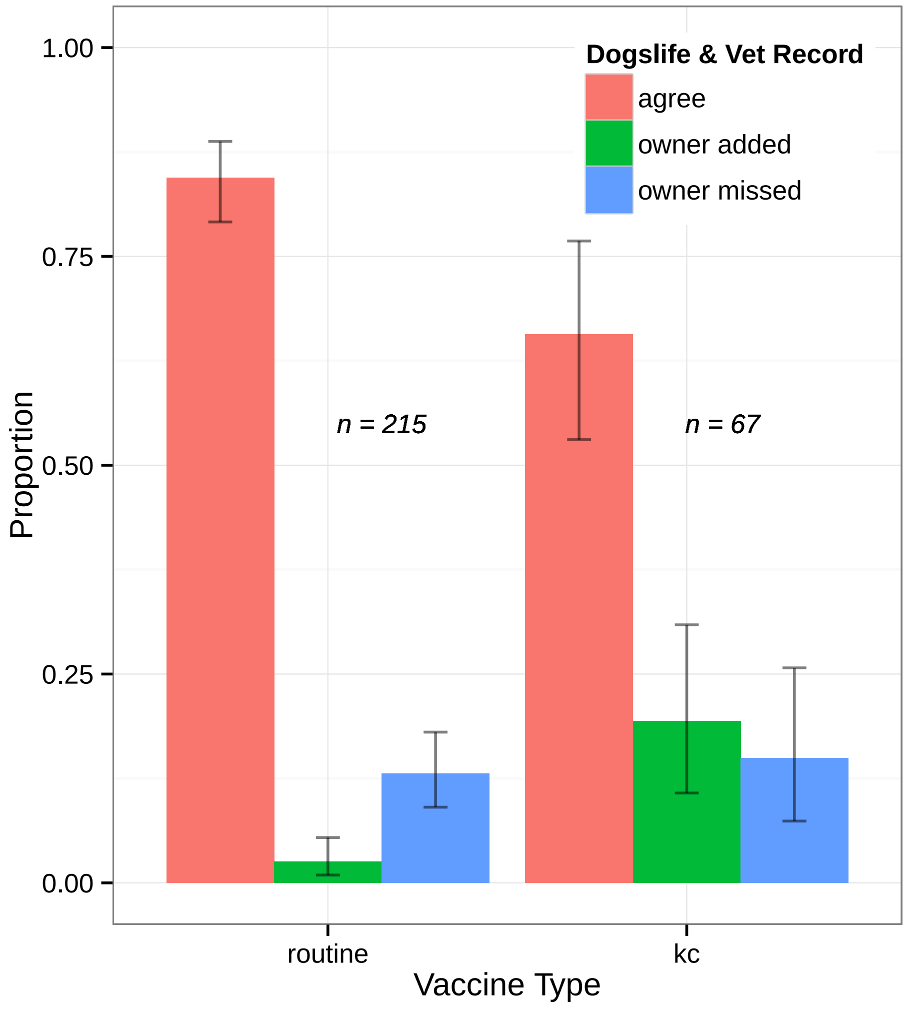


Figure 3.12: Proportion of vaccinations in Doglife records with exact 95% binomial CI. (kc = kennel cough)

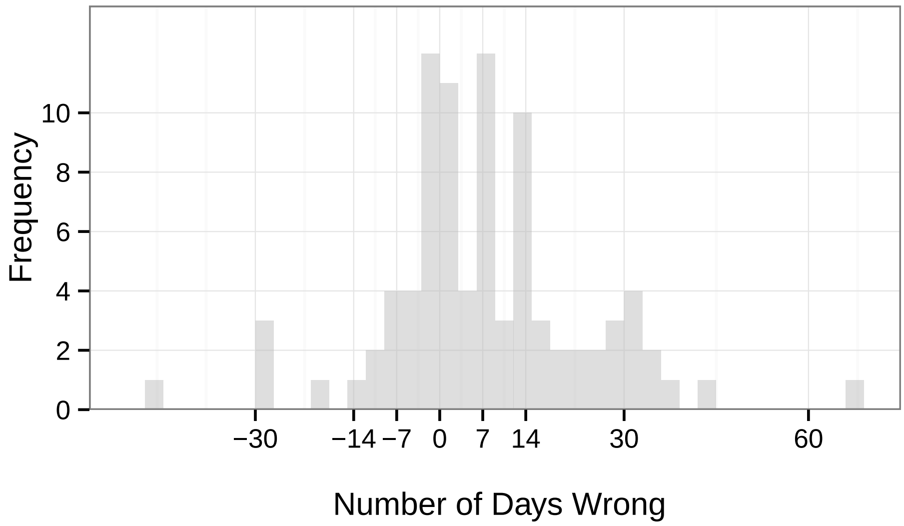


Figure 3.13: Frequency of vaccinations entered with incorrect dates

Of the 252 vaccinations that appear in both Dogslife and the vet records, only 163 were entered with the correct date. Figure 3.13 shows the distribution of the number of days 'wrong' with distinct peaks at ± one, + seven and + fourteen days ('out-by-day', 'out-by-week' and 'out-by-fortnight' errors). The majority are wrong by less than a month but one particular entry (omitted from the figure) was incorrect by 341 days and appeared to be an 'out-by-year' error.

To summarise, veterinary records for 128 dogs were examined and their Dogslife records indicated that 257 vaccinations were given. Of those 257, 236 were substantiated by the vet record (91.8%, 95% CI: 88.4 - 95.2%). A further 35 (13.6%, 95% CI: 9.4 - 17.8%) were found in the vet record but entirely omitted from the Dogslife record. The vaccination type did not appear to be related to the likelihood of that vaccination being omitted from Dogslife but the vaccinations found in Dogslife that were not substantiated by the vet record were more likely to be for kennel cough (13 of 57, 95% CI: 11.9 - 33.7%) than routine or rabies vaccinations (8 of 200, 95% CI: 1.3 - 6.7%).

Table 3.14: Number of veterinary visits and Dogslife 'illnesses'

		Veterinary Records	
		Yes	No
Dogslife Record	Yes	170	228
	No	440	unknown

3.4.4 Illnesses

As mentioned in Section 3.3.4.2, one page of the online questionnaire is titled 'Illness' (questions available in Appendix 2, Section A2.7). Owners are asked about six named presenting signs and 'Other' problems. They were initially asked about 'Scratching' and 'Licking and chewing' but after some confusion amongst owners, 'themselves' was added. i.e. 'Scratching themselves'.

If the owner states that they visited the vet for an illness, they are asked whether their vet filled in the Dogslife Veterinary Health Report. These reports comprise a side of A4 and one can be seen in Appendix 4. Owners are encouraged to print off blank copies when they first register. At the outset of the project it was hoped that owners would take a form with them each time they visited their vet and then use the expert answers provided to fill in identical questions online. Unfortunately the form is used in only 11% of cases. More typically owners do not use the form and instead give their perceptions of the illness from memory.

Non-routine veterinary presentations in the veterinary records and owner-perceived negative health events entered by the owner into the Dogslife database until the end of the veterinary record for 127 dogs were compared; this comprised 838 rows of data. Table 3.14 illustrates the level of agreement between non-routine veterinary presentations and owner reported illness events in the Dogslife database. There is no way of determining the number of illnesses that were neither in the veterinary records nor reported to Dogslife via the online questionnaire. Nor can the 228 health events reported by owners that did not precipitate a veterinary visit be validated. However 69% of these 228 health events were relatively unambiguous such as vomiting or diarrhoea so it can be hoped that it would be difficult to mistake these signs.

Initial comparison between the veterinary records and those reported health events associated with veterinary visits was disappointing (Table 3.14). Of the 610 veterinary visits, only 170 were reported to Dogslife. However the online ques-

Table 3.15: Number of unique illness events according to Dogslife and Veterinary Records

		Veterinary Records	
		Reported	Never Reported
Dogslife Record	Reported	137	205
	Never Reported	39	unknown

tionnaire was not designed to capture veterinary visits but rather clinical signs; enabling the capture of information (such as the 228 clinical signs mentioned above) regarding owner-perceived signs of ill health that are never presented at primary or secondary veterinary practices. Events that precipitated a veterinary visit would still be reported but the collection of a greater breadth of health information was facilitated. If an owner visited the vet three times for their dog's eye infection then unless the owner filled in the questionnaire on three separate occasions, they would not be given the opportunity to detail those three vet visits. The 610 veterinary visits included in this analysis referred to just 176 unique health events[1] and, as shown in Table 3.15, the Dogslife's record omitted entirely only 39 vet recorded health events - just 22%.

Health events which did and did not result in veterinary visits are compared in Figure 3.14. 'Scratching' and 'licking and chewing' have been collapsed into one category 'skin' and the two most highly represented areas in the 'Other' category have been separated into 'Other (ear)' and 'Other (eye)'. Gastrointestinal illnesses are disproportionately prevalent in non-vet visiting health events.

The health events which required a veterinary visit are shown in Figure 3.15 according to presenting sign and whether the owner reported them to Dogslife. The proportion of under-reporting appears uniform across health event type.

Figure 3.16 illustrates that the delay between an illness occurring and subsequent data entry is associated with the likelihood of the owner recording an illness via the questionnaire. For the illnesses found in the veterinary records, the median delay between a veterinary visit and when it was recorded with Dogslife was 16 days (IQR = 27.5 days). By contrast, vet visits that were omitted from the Dogslife record occurred a median of 40 days (IQR = 64.5 days) before the owner made a subsequent data entry.

[1] Grouping of veterinary visits performed by CP then checked by clinician Dylan Clements (DNC).

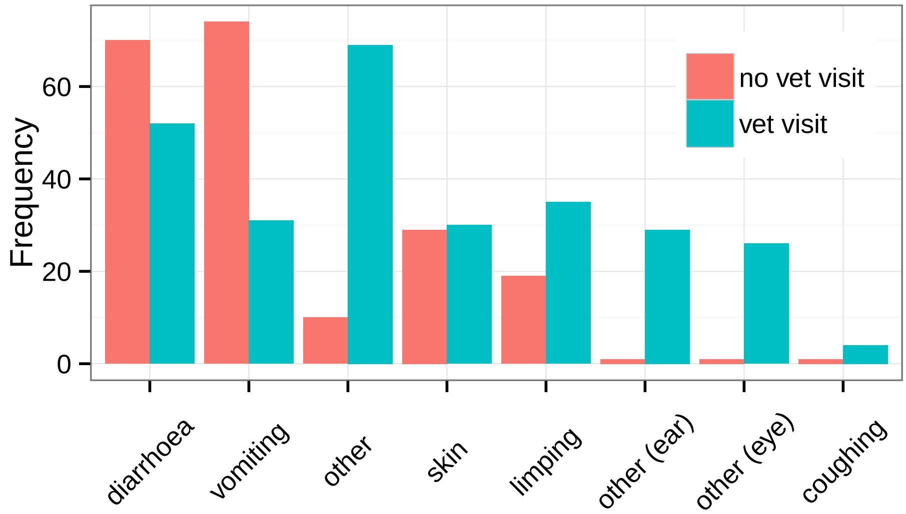


Figure 3.14: Health event type according to whether dog visited the vet

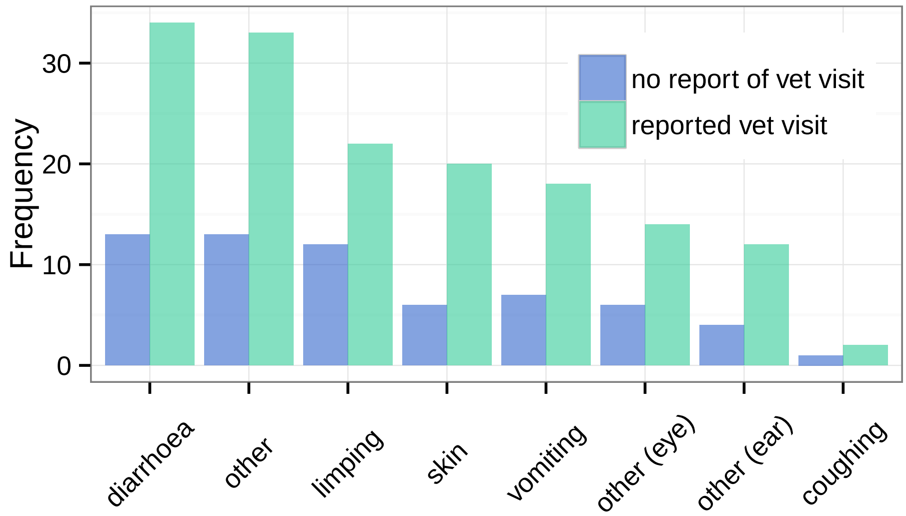


Figure 3.15: Health event type according to whether owner reported vet visit

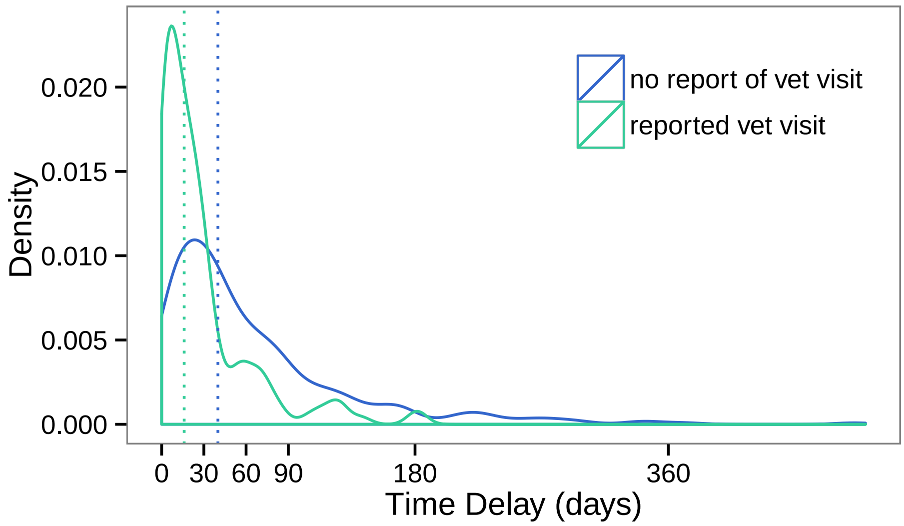


Figure 3.16: Time delay between vet visit and subsequent entry to Dogslife website. Dotted lines indicate median delay.

3.5 Discussion: Visits

3.5.1 Household Profile

Despite random selection of prospective households, Table 3.3 indicates that the owners visited were somewhat different to the Dogslife cohort as a whole. Only a small number of owners were involved so any differences between the sample and cohort such as the higher proportion of female dogs visited could be due to chance. However the nature of the household types visited appears non-representative of the cohort and likely reflects responder bias. Despite offering to visit during evenings and at weekends in addition to normal office hours, a visit still meant a time imposition for owners. It might be expected that retirees would have more free time and therefore be over-represented in this group. That this same over-representation is repeated for the singleton households is less easy to explain but may again reflect a higher amount of free-time in this group relative to the more populous households.

In contrast to the Dogslife visit responder bias, studies of human health typically find their response bias to have a different slant. Direct comparisons between responders in surveys of human health and those visited by Dogslife are problematic because the visited owners were sampled from a Dogslife cohort that has effectively been pre-screened for interest in Labrador Retriever health. Instead humans who volunteer for health studies might more properly be compared with owners who originally join Dogslife. Nevertheless, a study of human health in the Netherlands found that households containing only one person were *less* likely to respond to the survey (Reijneveld and Stronks [1999]). Indeed whilst the Dogslife household categories are not directly comparable with much of the published literature, married people (approximating to the 'Family' and 'More than one adult' categories) are normally *over*-represented as respondents in health surveys. Age is a more complicated factor and its effect on response appears to depend on the survey. The millennium cohort found that old age was associated with higher response (Littman et al. [2010]) but the Netherlands study found no effect (Reijneveld and Stronks [1999]). Owners who did not volunteer to be visited were not asked for their reasons but it should be recognised that they may represent a different group of dogs than those examined.

3.5.1.1 Household Type

Despite the effort to define the household options clearly, the categories were not mutually exclusive and some degree of confusion regarding classification was inevitable. For example households might include two adults but only one might be retired, placing this household into two Dogslife categories. Using titles such as 'Family' also allowed owners to use their own definitions. In particular when visiting one couple it was apparent that the presence of children was not relevant to them when calling themselves a family. With hindsight, removing 'Family' and specifying the age of children by re-titling the first category as 'One or more adult and one or more children under 16' might help owners choose categories. Similarly changing the 'Retired' category to 'One or more retired adults' might be clearer. In terms of data usability, the misclassification was minor and should have little effect on future analyses.

3.5.1.2 Smoking

The visits highlighted a potential issue of under-reporting of smoker status amongst those who only smoke outside. This may in part be due to the phrasing of the question "Does anybody in the household smoke?" which may have been interpreted as "Does anybody smoke in the household/house?". Self-reporting of smoking status is thought to be reliable in the general population (Patrick et al. [1994]) but affected by social desirability in specific groups such as pregnant women (Russell et al. [2004]) and those suffering from respiratory diseases (Lewis et al. [2003]). Smoking is increasingly socially undesirable in the UK and this may also have contributed to the under-reporting of smoking in the Dogslife cohort.

There is no way to quantify the impact of misclassification of outdoor smokers as non-smokers but a recent study in Scotland would suggest that the effect on their dog may be minimal (Knottenbelt et al. [2012]). The investigation was not primarily focussed on indoor versus outdoor exposure but compared three groups of dogs: those never exposed to cigarette smoke, those exposed irregularly or outdoors, and those exposed regularly indoors. The levels of nicotine found in the neck fur of dogs in the never exposed and irregularly exposed groups were clearly distinguishable from the regularly exposed group. As might be expected, those in the group that included dogs only exposed outdoors had intermediate levels of nicotine (between no and regular reported exposure) but analysis indicated that there was no evidence of a difference between the never exposed and irregularly exposed groups. From a Dogslife perspective, this negative result would make the mis-classification of outdoor smokers as non-smoker households unimportant in future analyses.

3.5.1.3 Dog Purpose, Other Pets & Geographical Location

Visited owners appeared to find the 'dog purpose' and 'other pets' elements of the online questionnaire simple and easy to answer. None of the owners suggested amendments and their visit answers broadly agreed with the answers previously given online. Owners are also asked for their postcode at registration and of the 43 households visited, one owner was visited at a postcode that was different to the location given in the Dogslife database. The owner mentioned having moved in the previous year but they had not updated their profile.

3.5.2 Repeated Dogslife Questionnaire

3.5.2.1 Sleeping Location

There was broad agreement between visit and Dogslife database answers regarding the dogs' sleeping locations with 36 of 43 owners giving the same answer in both instances. Eleven of the visited dogs slept with another dog and their owners categorised them either as 'sleeping alone' or 'other'. Following the visits, amendments were made to the potential answers to "Where does Dog* sleep at night?" so that owners may now explicitly choose an option of 'sleeping with another pet'. With this amendment, 40 of 43 owner answers would agree.

3.5.2.2 Diet

The Dogslife questionnaire attempts to capture daily, repeated food intake by asking owners to weigh their dogs' meals. These weight measurements were relatively reliable for wet food but were of varying quality for the other food types. In some cases, there appeared to be a problem with owners giving the weight of a meal rather than the whole daily quantity. Whilst visiting it also became apparent that the full range of food eaten by the cohort could not be captured by the Dogslife questionnaire and that, despite offering examples online, owners did not necessarily agree on what might be considered a titbit. Where owners did mention feeding titbits, they varied from dog training treats to nightly bowls of rice pudding. That type of detail is not captured by the database and likely could not be usefully quantified given the number of Dogslife participants but its absence further increases the difficulty of comparing diet and calorific intake based on the online questionnaire alone.

Whilst it is impossible to know why owners found the "titbit" question so difficult to answer accurately, the use of the word titbit may be part of the problem. There may be an assumption in owners minds that a titbit comprises human food, from a human's plate; or perhaps the word titbit implies food that is inherently bad for their dog. Upon reflection, the word titbit should perhaps have been avoided. In light of the apparent difficulty owners had answering the titbit question, it was re-phrased as follows:

"Does Dog* also receive"titbits"? For example anything else your dog eats such as food off your plate, training treats, chews etc.?"

This new phrasing did not remove the word titbit but the answers given before and after the phrasing change indicated that owners apparently adapted their answers. During October 2011, there were 410 entries regarding the titbit question comprising 245 no and 165 yes answers. In October of 2012, following the wording change, there were 499 answers comprising 85 no and 414 yes answers. Without visiting the 85 owners who answered no and asking them face-to-face it is unclear whether they too fed their dog "just a little bit of chicken skin now and again" or truly did not give their dog anything extra, but the new phrasing does yield answers which are closer by proportion to those found during the visits.

Unfortunately it appears that the data captured by the online questionnaire are not accurate at a fine level of dietary detail. Nevertheless the data provided by owners are valuable at a level of broad discrimination such as whether the owner feeds a primarily wet or dry diet.

3.5.2.3 Bathing

The phrasing of the bathing question when Dogslife was launched ("Has Dog* been bathed in the last 4 weeks?") presented a problem because after the dogs reach one year of age, owners are only asked to make data entries every three months. In the light of that, an owner might have said that their dog had been bathed during the visit but it would legitimately never appear on the Dogslife database. The visit process highlighted this issue and the question was changed to "Has Dog* been bathed since you last visited the site?"

The most frequent answer to how often the dog was bathed was 'When (s)he needs it' and despite prompting owners were not always able to turn this into an answer allowed by the Dogslife questionnaire. The visited owners did not bathe their dogs on a regular basis so for them, the potential answers to the bathing frequency question were inappropriate. Prior testing of the Dogslife questionnaire did not highlight this issue, presumably because for the owners surveyed, the offered answers were not 'wrong' but rather 'not ideal'.

Despite these issues, when owners had never made a bathing entry into the database, the visit answers indicated that it was reasonable to assume that the dog had likely never been bathed. Thirteen of fifteen owners asserted in person that their dog had not been bathed and the remaining two had only been bathed once. Changing the phrasing of the initial bathing question should have reduced the number of 'Never' misclassifications in the data. Typically when an owner had entered 'Less than monthly' their answer was also substantiated in person. Only two of thirteen database answers of 'Less than monthly' were contradicted during the visits. The monthly and fortnightly measures seemed to be less reliable and the more frequent options were not tested as none of the owners visited bathed their dog so frequently.

3.5.2.4 Swimming

The online questionnaire does not address the question of dogs paddling and/or swimming. During the visits owners were asked whether their dogs accessed water in order to gauge whether such access need be considered as an environmental risk factor in future analyses. The addition of a paddling/swimming question to the online questionnaire was thought to be too onerous for owners to address in each data entry but 35 of the 44 dogs engaged in swimming or paddling to varying extents. It was therefore deemed important to ascertain whether a dog had access to water when considering specific diagnoses such as otitis externa and limber tail. Investigations into either condition could be facilitated by one-off questionnaires sent to smaller sub-samples of the cohort without repetitively burdening owners.

3.5.2.5 Weight

Weight is the one question in the online questionnaire that is not mandatory but many owners mentioned making special trips to the vet in order to weigh their dogs for the project. Despite some concern on the part of the project team that dog weight might be particularly hard for Dogslife to capture (due to lack of data), many owners offer dog weights and the database weights of those visited appear reliable. This reliability could not be directly tested for dogs under one year during the visits as the dogs were still growing. As might be predicted, all but one of the visit weights were higher than those in the Dogslife database. However where weights were mentioned in the veterinary records, they were fully consistent with the Dogslife records. Comparison between database and visit weights for dogs over one year was more legitimate and there was a high level of concordance between the two measures. As the cohort age, weight reporting may become subject to reporting bias with owners of obese dogs more likely to not report their dog's weight due to social stigma but to date, Dogslife recorded weights appear to be a good reflection of the weight of the cohort.

3.5.2.6 Height

Whilst it might be expected that the agreement between Dogslife and visit height measurements would be poor for the younger dogs, it was also poor for dogs over one year. A video is available on the Dogslife website showing a standard method for measuring dog height but many owners commented on the difficulties they faced when obtaining this measurement. Indeed the dog measurement page (which includes a mandatory height measurement question) has the highest drop-out rate of any page of the online questionnaire (Clements et al. [2013]). Owner (and visitor) problems when taking the measurements will likely be reflected in increased noise in the height data but these difficulties cannot entirely explain the *shape* of the disagreement, particularly for the older dogs. As expected delays between online reporting and visit measurement had a greater effect on comparisons of younger dogs but the effect was still present in older dogs. This clear effect on the older age group indicated that the stadiometer was unlikely to be the cause of measurement difference between the database and visit measurements. Not only was the agreement minimal, but only four of the 29 visit measurements were lower than those entered into the database. If it is assumed that the stadiometer is equivalent to the methods used by owners and that the dogs stop growing at one year, then the measurements taken during the visits should have a 50% chance of being lower than those entered by owners i.e. 14-15 measurements should have been lower. The lack of balance in Figure 3.8 suggests that either the stadiometer measurements were inaccurate, the database measurements used were inappropriate (due to the relative youth of some dogs) or that the dogs continued to grow after one year of age.

During 2012, two owners contacted Dogslife after their dog's first birthday to indicate that their dog had grown but that they were no longer able to enter height measurements because their dog was over one year of age. Subsequent literature searches found a new study regarding skeletal growth in Labrador Retrievers (McBrien et al. [2011]). Tantalum markers inserted into the tibia of six male dogs were shown to still be separating as the study concluded when the dogs reached 13 months of age indicating that the dogs had not stopped growing. On the basis that it was apparently unclear when Labradors reached full height, the website questionnaire was modified in early 2013 so that owners were asked to enter their dog's height until the dog was 18 months old. It was anticipated that this extension would potentially help solve another issue in the height data that was highlighted by the visits - the age at last height entry of the older visited dogs in this analysis was often well under one year. By offering the opportunity to enter height data after one year, owners were given the chance to 'catch up' and give Dogslife the height of their mature dog.

3.5.2.7 Visit Summary

At the time of the visits, information given as a one-off at registration was typically repeated. Owner characterisation of their dog's purpose and answers regarding other pets in the household were largely in agreement with those previously entered into the Dogslife database. Only one owner had moved house but not revised their online profile and a different owner had stopped smoking but not updated their smoking status. Under-reporting of smoking status by those who smoked outside was an issue in this sub-section of the cohort but recent work suggests that this under-reporting should not greatly affect any future risk factor analyses (Knottenbelt et al. [2012]). Collection of household type information was complicated by non-mutually exclusive categories but even with this issue, 37 of the 43 visited households were accurately characterised and up to date in the Dogslife database.

As the examples of the owner who moved and the owner who stopped smoking indicate, irrespective of reminders, not all owners will update their Dogslife profiles. However one owner explicitly mentioned doing so when one of their other dogs had died and the facility helps minimise inaccuracy due to changes over time.

The visits provided a great deal of useful feedback from owners regarding the project. Several owners commented that they would like to be able to interact with other owners and compare their dog's progress with the rest of the cohort. There was also a lack of enthusiasm regarding the repetitive nature of data entry; particularly when nothing was changing. For example owners are asked each time what weight of food they give their dog and for the details of the dogs' exercise regimes. The answer fields were not pre-populated with the previous answer in the hope that owners would revaluate their answers each time and Dogslife would capture any subtle changes. It became apparent during the visits, particularly for older dogs, that both diet and exercise regimes did not vary and that owners were not weighing the food or re-estimating their dog's exercise. Retention becomes an increasing priority as a cohort ages so addressing the concerns expressed during the visits was a high priority.

In the interests of preventing the cohort from influencing each other and biasing all results, owners could not be facilitated in directly comparing their dogs. However in Janurary 2013 the first paper based on the Dogslife project was published (Clements et al. [2013]) and the study-so-far section of the website was updated to feed these results back to the owners. In November 2013, sleeping location, dietary answers and exercise answers were pre-populated in the online questionnaire for dogs over one year of age in the hope that it would minimise owner response fatigue and maximise retention. Mooney et al. [2009] addressed the issue of pre-populating factual answer fields as part of their longitudinal study of substance abuse. They showed that offering partially pre-populated questionnaires resulted in minimal time saved for the substance abuse centres who completed them but was perceived to be "very helpful" - a positive effect on perception rather than reality. Given that Dogslife is based on the goodwill of owners, such a perceived impact was thought to be invaluable in maintaining data entry levels.

3.6 Discussion: Veterinary Records

3.6.1 Prophylactic worm and flea treatment and neuter status

Comparing the worm and flea treatment information in the vet records with that in the Dogslife record was extremely difficult. The quality of data available in the vet records made validation of the Dogslife record largely impossible. Neuter status however was well recorded in both the vet record and in Dogslife. The two records were very much in agreement and neuter status and timing in the Dogslife record can be relied upon in future investigations.

3.6.2 Vaccinations: Data Accuracy

Vaccination reporting was not entirely accurate. Owners reported vaccinations that could not be substantiated from the veterinary records, failed to report others included in the records and only 163 of 252 reports had correct vaccination dates. In particular, owners were likely to report that their dog had received a kennel cough vaccination when they had not. This last finding perhaps suggests that owners believe kennel cough is vaccinated against as part of their dog's normal yearly booster. In future analyses, kennel cough will be the least reliable of the vaccination measures captured by Dogslife. However, despite some misclassification of vaccination status, the vaccination reports give a guide to routine health care for the dogs. Rabies and routine vaccinations appear broadly reliable with under 20% of vaccination courses missed entirely.

3.6.3 Illnesses: Data Accuracy

Veterinary and Dogslife records were examined for 127 dogs; 381 different illness events were recorded in one or both media. Of the events that involved a veterinary visit, only 39 (22%, 95% CI: 16.3 - 29.0%) were entirely omitted from the Dogslife record and the proportion of each type of event reported lay uniformly between 75% - 80%. Events which required a vet visit that were reported to Dogslife were accurately described by owners resulting in Dogslife presenting signs that were consistent with the veterinary records. Gratifyingly, the Dogslife diagnosis records were also consistent with the veterinary records. In future risk factor studies, the Dogslife records maybe confidently used to accurately classify vet-visiting cases.

The majority of health events (205) did not involve a veterinary visit and these conditions were disproportionately distributed between the Dogslife categories. Illnesses involving vomiting and/or diarrhoea were more likely not to involve a vet visit, which is consistent with surveys of human vomiting and diarrhoea such as a study of infectious intestinal disease in the UK that found only one in six cases were reported to a doctor (Wheeler et al. [1999]). By contrast, conditions involving limping or lameness, or 'other' were more likely to involve a veterinary presentation than not, whereas coughing, licking or chewing, and scratching were approximately equal in terms of numbers visiting and not visiting the vet. The data regarding these non-vet-visiting illnesses is of unknown quality, both due to lack of validation and because all detail relies on owner reporting in the absence of veterinary diagnosis. As such it is typical for these health events to have been coded within Dogslife as lacking a diagnosis. However, their presenting signs were recorded and these presenting signs may be used as case definitions in future analyses. Much of any potential unreliability in non-vet-visiting illnesses is inevitable by definition but these data should prove invaluable for building a picture of a dog's ongoing health.

3.6.4 Vaccinations & Illnesses: Recall Decay

The delay between vaccination and subsequent Dogslife data entry was clearly associated with vaccination reporting. Omitted vaccinations were associated with a greater delay. This phenomenon, known as recall decay, can be explained by a theory of declining *strength* of a memory (Hinrichs [1970]). It is taken into account in the design of censuses (Neter and Waksberg [1965]), specifically studied with regard to the accuracy of self-reported drug misuse (Shillington and Reed [2011]) and must be considered in the design of all questionnaires. The design of Dogslife as a web-based project should have minimised recall decay as owners may enter new information at their convenience and are not limited to reporting when an interviewer visits or the next questionnaire comes through the post. However the interval between Dogslife data entries for the young cohort, discussed in Clements et al. [2013], peaks at 37 days when owners receive a reminder email and more recent analysis shows a largely bimodal distribution of intervals for the whole cohort with a peak for owners of dogs over one year at approximately 97 days. As such, the delay between an incident and the subsequent reminder email introduces the potential for recall decay.

In the case of the Dogslife vaccination record, the vaccination recall decay plot (Figure 3.11) has a remarkably similar shape to the illness recall decay plot (Figure 3.16) indicating a similar pattern of recall amongst the owners for both questions. The median delays for reported and unreported incidents are longer for the vaccinations but this may reflect the inclusion of vaccinations that took place prior to Dogslife registration rather than any true difference in recall period. In order to minimise misclassification of dogs as controls when they should be cases (or unvaccinated when they were vaccinated), use of the vaccination and illness data in Dogslife may involve choosing 'at-risk' periods of time for each dog that encompasses a number of days prior to each data entry rather than considering their whole record. Schmidt et al. [2007] used this technique in a longitudinal study of diarrhoea prevalence and while it reduced the quantity of available data, quality was improved. Feikin et al. [2010] used data from 53,000 people in Kenya to determine that reported disease prevalence dropped below 80% of the maximum prevalence when people were interviewed more than three days after an incident for children under five years of age and more than four days after for individuals over five years of age. Such precise estimates of the effect of recall decay cannot be made with a relatively small number of vet records but using a limited 'at-risk' period and, for specific studies of high prevalence diseases, potentially requesting further vet records to ensure case and non-case status will increase the power of risk factor studies using Dogslife dogs.

Steps have also been taken to encourage owners to return when an incident happens rather than when their reminder email arrives. Firstly the pre-population of data in the questionnaire is aimed to reduce questionnaire fatigue so that owners are not put off from coming to the Dogslife website and may return more frequently. More directly, owners may also update their dog's illness information without completing the full questionnaire. Once an owner logs into their profile on Dogslife there is a link, either on their homepage or visible on the "Lab report!" health page, saying "If you need to update information about an illness, you can **update *Dog's** illness information**." It was hoped that this would encourage and facilitate owners with regard to contemporaneously telling Dogslife about any illness. It is not possible to determine how many owners make use of this link but following the validation process, it was decided to advertise the link as the first line in each of the monthly Dogslife newsletters. There is a peak in website traffic following each newsletter (Clements et al. [2013]) suggesting that it has a positive impact on cohort retention and it is hoped that advertising the health link in this fashion will increase its use and minimise recall decay.

3.7 Conclusions

The initial aim of the visit process was to assess the construct validity of the Dogslife questionnaire. With the exception of the dog measurements, it was not a traditional validation process as Dogslife data were not compared with an independent gold standard. Instead database answers were checked against answers provided by the same source of information at a later date - agreement was found to be good. It was an invaluable process and it became clear during the visits that the 'comments' sections of the visit questionnaire held the key to improving the online Dogslife questionnaire. Several minor changes were made to the answer options throughout the questionnaire and a major decision was made to pre-populate the more time-intensive questions for dogs aged over one year. It was hoped that the amendments would emphasise Dogslife willingness to respond to owner feedback and facilitate retention of the cohort.

Comparison of the veterinary records with Dogslife records led to the implementation of efforts to minimise recall decay. A decision was also taken to minimise the effect of recall decay on potential misclassification bias by using data from limited periods of time prior to each data entry rather than the entirety of each dog's record. With the exception of a proportion of the kennel cough vaccinations, where data were available, Dogslife vaccination records were found to agree with the vet records and as a whole the Dogslife database offers a remarkable resource for future investigations of Labrador Retriever health in the UK.

Chapter 4

Lifestyle & Morphology

4.1 Introduction

At midnight on 31st December 2013, a copy of the Dogslife database was captured. It comprised data collected over a period of three years and six months for dogs that were aged up to four years. There is a dearth of published, peer-reviewed work regarding what is normal for UK dogs and the Dogslife project was designed to address such knowledge gaps. The aim of this chapter is to describe the morphology and lifestyle of Dogslife dogs using Dogslife data. A condensed version of this chapter has been published by Pugh et al. [2015b].

4.2 Methods

As mentioned in Section 2.3.1, raw data from the database were cleaned into separate csv files. These csv files were used in all future analyses and it was possible to link them together by the use of unique dog, owner, data entry or illness instance identifiers (IDs). The different methods used for data cleaning are described in association with specific data types.

Appropriate statistical techniques described in Section 2.3.1 were used to seek associations between demographic factors. Linear mixed models were performed using the *lme4* (Bates et al. [2014]) and *nlme* (Pinheiro et al. [2013]) packages. Model assumptions regarding the residuals (normality and homogeneity of variance) were checked by visual inspection of plots of residuals against fitted values.

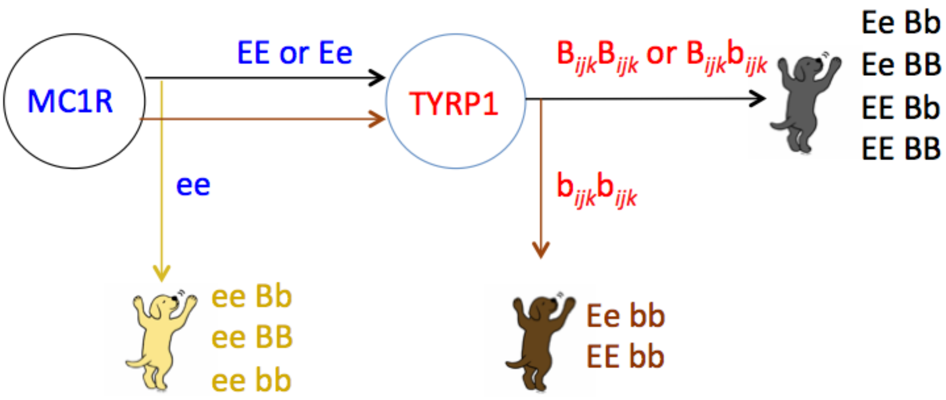


Figure 4.1: Genetics of common Labrador Retriever coat colours

4.2.1 Coat colour analyses

Coat colour genetics form part of analyses included below. Saliva samples sent by owners (as described in Section 2.2.4) were genotyped for coat colour by Ailsa Carlisle, Roslin Institute. Polymerase Chain Reaction (PCR) based amplification of *MC1R* and *TYRP1* was undertaken using methods which are standard in the laboratory of Professor Kim Summers. In LRs, yellow, black and chocolate colours are controlled by two genes *MC1R* and *TYRP1* (Schmutz and Berryere [2007]). At *MC1R*, a homozygous mutation (ee) characterised by a stop codon instead of an arginine results in yellow coat colour irrespective of genotype at *TYRP1* (fox red dogs are a variant of yellow so they are also ee at *MC1R*). *TYRP1* controls the production of a tyrosinase-related protein which differentiates between chocolate and black dogs should *MC1R* be heterozygous (eE) or homozygous wild type (EE) (Figure 4.1). Three different recessive mutations at *TYRP1* can result in chocolate dogs (bb).

4.2.2 Time at risk

The cohort time at risk comprised the sum of each dog's individual time at risk. This was looked at in two ways. The first time an owner filled in a questionnaire, all questions asked about the preceding four weeks. The total time at risk was therefore considered to start 28 days before the first complete data entry and end at the final completed data entry. If the first data entry took place before the dog was 28 days old (4 dogs), the date of birth was used as the starting point.

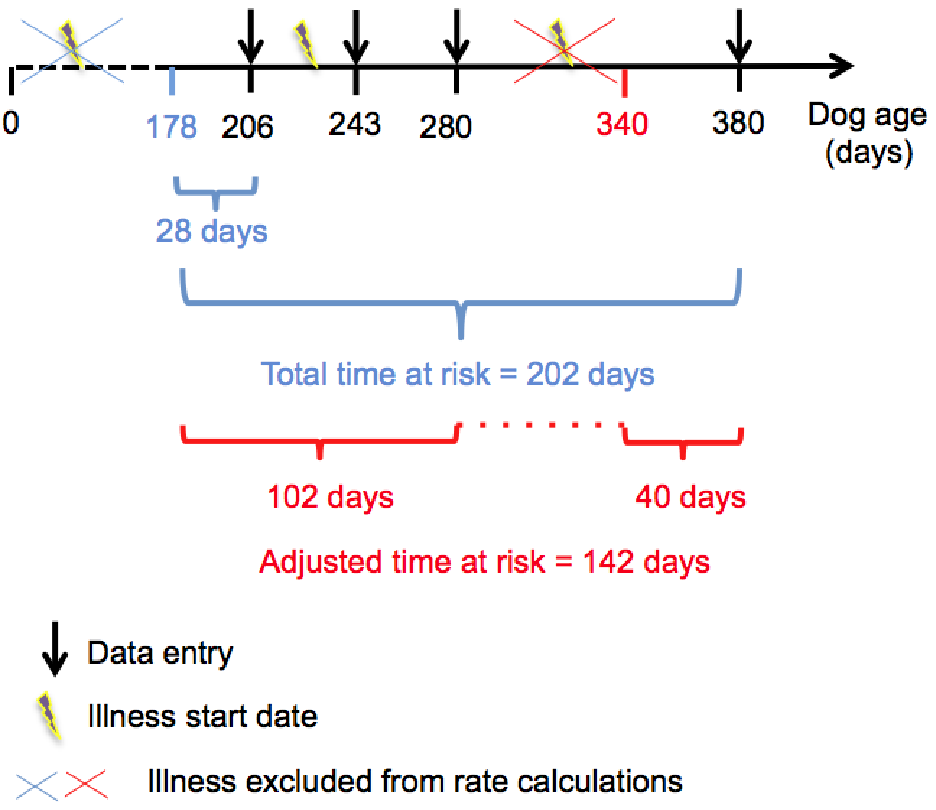


Figure 4.2: Calculation of total and adjusted times at risk. The dog was registered at 206 days of age and windows of 40 days prior to data entries were used to calculate adjusted time at risk.

For subsequent data entries, there was evidence of recall decay when owners left long gaps between Dogslife visits (Section 3.6.4). Adjusted times at risk were therefore calculated considering a fixed period prior to each data entry. The impact of different time windows was explored. Figure 4.2 shows the calculation of an adjusted time at risk using a 40 day window. Illnesses and vaccinations that were reported to have taken place more than 28 days before the first data entry or more than, for example, 40 days before the entry when it was reported would be included in frequency reports but excluded from rate calculations.

The dog in Figure 4.2 would contribute 202 days to the total cohort time at risk but just 142 days to the time at risk if the window used was 40 days. They would be considered to have suffered three illnesses but only the latter two would contribute to illness rate estimates that addressed the total time at risk. This would reduce to just one illness (the middle one) if a 40 day window for time at risk were applied.

4.2.3 Relatedness

The sires and dams of the members of the Dogslife cohort were tabulated using their KCID in order to determine how many pups each sire and dam contributed to the cohort. The table contents were then themselves counted to give a measure of how many dams and sires had one, two, three etcetera pups in the Dogslife cohort. The frequencies of sire and dam contributions in the cohort were then compared with the the same statistics generated by 1,000 cohort-sized groups that were randomly selected from all eligible dogs.

4.2.4 Dog Height Cleaning & Analysis

Owners were asked to measure their dog's height to the shoulder (as described in Section 3.2.1.1; question available in Appendix 2, Section A2.2). The website asked for a two digit number and gave the owner the opportunity to choose the appropriate units (inches or cm). If the owner chose inches, their entry was converted by multiplying by 2.54 and stored in the database in cm. The question was asked for dogs aged under 18 months and asked again when each dog reached three years. An answer must have been entered for the owner to continue on through the questionnaire. All raw reported heights were plotted against dog age (Figure 4.20) and it was apparent that extensive cleaning would be required. The steps of the cleaning process are described below.

Step 1: At times, owners appeared to try to correct their own data entry errors. For example, one owner entered 254 cm, then within the same day entered 24 inches (60.96 cm). The data were reduced such that if entries were made within five days of each other, only the latest height would be kept.

Step 2: Unusually high heights were examined in the context of the height records for affected dogs. Measurements such as 81.28 cm (entered as 32 inches) were assumed to be the result of typographical errors whereby the owner should have entered 23 inches (58.42 cm). Amendments were made where it was consistent with the dogs' records.

Step 3: Heights of less than 7 cm when the dog was aged over 50 days were examined in the context of the height record for that dog. The assumption was made that, should they not be the result of obvious typographical errors, then perhaps the owner had entered a false number in order to continue past the compulsory height question. All 'false' entries of 1 cm or less were deleted. If the dog was aged between 50 and 100 days, all 'false' entries of 5 cm or less were deleted. If the dog was aged over 100 days, 'false' entries of 7 cm or less were deleted. These boundaries were chosen on the basis that the heights were extreme outliers for those ages.

Step 4: Figure 4.22 shows what appear to be two growth curves. The lower curve is approximately 2.54 times lower than the main curve. It was hypothesised that owners had taken a measurement in inches but chosen cm as the units when entering the record into Dogslife. It was also thought possible that some of the very high heights might be the result of owners measuring in cm but entering their dog's height in inches. To correct these heights, a probabilistic model was applied to the whole data set and used to estimate whether entries might have been given in the correct or incorrect units. The heights were assumed to be normally distributed (Equation 4.1a) and the mean height was modelled to change exponentially with age (Equation 4.1b). Each height would also fit one of three classes (Equation 4.1c): measured in cm and reported in inches; measured and reported in the same units; measured in inches and reported in cm.

$$height = \mathcal{N}(\mu, \tau) \tag{4.1a}$$

$$\mu_i = A(1 - e^{(-B(age_i - C))}) * class_i \tag{4.1b}$$

$$class = \begin{pmatrix} \frac{1}{2.54} \\ 1 \\ 2.54 \end{pmatrix} \tag{4.1c}$$

The model was estimated under a Bayesian framework using the *rjags* package (Plummer and Stukalov [2014]). Each sex was modelled separately. One thousand iterations were used for adaptation and 2,000 were discarded as 'burn-in'. The final model was based on a further 5,000 iterations and the mixing of the models was checked using the *coda* package (Plummer et al. [2006]). Once identified, the mis-reported heights were corrected using a multiplier of 2.54 or $\frac{1}{2.54}$ and these corrected heights were used in future analyses.

The model required Bayesian priors, shown in Equations 4.2. Parameter A was the mean full height of the dogs (Equation 4.2a) and was taken from the UKKC breed standard for LRs which was 55-56 cm for females and 56-57 cm for males (The Kennel Club [2014]). Parameter B was a proxy for growth rate

(Equation 4.2b). The height was 'decaying'; growing half way closer towards its maximum height, A, every $\frac{ln2}{B}$ days. Parameter C was an off-set term that allowed the height to have a non-zero value when the pups were born (Equation 4.2c). Parameter *pi* is the prior probability of a measurement belonging to each different error group (Equation 4.2f): i.e. estimated 10% chance of being subject to each type of inches-cm error and 80% chance of having the correct units.

Bayesian Priors:

$$A = \mathcal{N}(56, 0.01) \tag{4.2a}$$

$$B = Uniform(0, 1.5) \tag{4.2b}$$

$$C = Uniform(0, 100) \tag{4.2c}$$

$$\tau = Gamma(0.001, 0.001) \tag{4.2d}$$

$$sd = \sqrt{\frac{1}{\tau}} \tag{4.2e}$$

$$pi = Dirichlet(0.1, 0.8, 0.1) \tag{4.2f}$$

Step 5: Following correction, there remained heights greater than or equal to 75 cm. These were all examined in the context of the relevant dog's height record. Heights that were assumed to have been over-corrected by the Bayesian approach were amended by dividing by 2.54. Heights that appeared to be affected by mis-typing, such as that described in Step 2, were amended. Heights that could not be simply amended were deleted.

Dog heights were then examined with reference to sex and coat colour using the Bayesian approach described above. Mature heights were compared with UKKC breed standards.

4.2.5 Dog Weights Cleaning

Owners were asked for the weight of their dog regardless of its age but they were not required to answer the question (as described in Section 3.2.1.1; question available in Appendix 2, Section A2.2). They were given the opportunity to choose their unit of measurement and if they entered the number in poundss (lbs), then it would be converted to kg by dividing by 2.204623. As with the heights, the dogs' weights were plotted against age and examined visually (Figure 4.26). Cleaning was undertaken in a stepwise process that is described below.

Step 1: As with Step 1 of the cleaning process for heights, owners appeared to correct their own data entry mistakes. If multiple entries were made within a five-day interval, only the last weight was kept.

Step 2: Weights above 100 kg were examined in the context of the weight record for that dog. Measurements such as 3,705 kg were changed to 37.05 kg. Similarly, weights such as 269 kg were changed to 26.9 kg.

Step 3: Weights of less than 17 kg when the dog was aged over 300 days, less than 12 kg when over 200 days or less than 5 kg when over 100 days were examined in the context of the weight record for that dog. Weights were amended if they appeared to be subject to obvious typographical error. Individual weights were deleted if they could not be reconciled with the rest of that dog's record. If the whole record seemed low then the height of the dog was checked. Dogslife records were searched for pictures of dogs that seemed to be outliers. If no further information could be found, these measurements were assumed to be correct. Individual measurements that appeared to be low by a factor of ~2.2 were assumed to be the result of a lb-kg error, similar to the cm-inch error for heights. These weights were multiplied by 2.204623.

Step 4: All weights over 50 kg were examined in the context of the dog's complete weight record and any photographic information. As for the low weights in Step 3, typographical errors were amended. Inexplicably high weights were deleted. Those that appeared to be subject to a lb-kg error were divided by ~2.2 and those that fit with an increasing overall weight profile for a dog were left.

4.2.6 Dog Weights Analysis

Dog weights data were explored visually with reference to coat colour, dog sex and other non-varying factors such as dog purpose. Weight was used as a proxy for an investigation of obesity using only weights when the dogs were aged one year and over. A linear mixed model approach was used. Both pet and owner IDs were included as random terms and an auto-correlation structure was used to account for the repeated measurements. Univariable analyses were undertaken before combining factors into a model to describe weight variation. Unchanging factors were initially examined before moving to include variables that changed with time such as sleeping location, neuter status, diet and exercise levels. Exercise levels were investigated as both categorical and continuous variables. The focus of the model was on main effects but biologically plausible interactions between age, sex, neuter status and height were also assessed. The reported multivariable model had the lowest Akaike Information Criterion (AIC) of all possible models, found using the MuMIN package (Bartoń [2014]).

Height needed to be included in the model but finding a reliable, mature height for as many dogs as possible was complicated. As was demonstrated during the validation process, individual height measurements were not very reliable and it was not clear when the dogs stopped growing (Section 3.3.4). A compromise was sought between including as many dogs as possible, many of whom might not have heights reported as they got older, and trying to ensure that the heights used were mature heights. It was also not thought advisable to simply use the tallest height for each dog after they reached a specific minimum age because the reliability of individual measurements might bias the outcome. Only heights reported after the dog was 10 months of age were considered and for each dog, the mean of these older heights was taken.

The results of the mixed model showed an association between coat colour and weight such that chocolate dogs were heavier than their yellow and black counterparts. It was assumed that a proportion of yellow dogs would have chocolate mutations and that they would effectively be "stealth" chocolate dogs (Figure 4.1, genotype eebb). If the relative heaviness of the chocolate dogs were associated with the chocolate mutation in *TYRP1* then it was hypothesised that heavier yellow dogs might have the chocolate genotype. DNA samples from 25 yellow dogs were chosen according to sex and weight. Canine *TYRP1* was amplified using PCR techniques by Ailsa Carlisle, Roslin Institute, University of Edinburgh (Section 4.1). The PCR products were purified and amplified before being sequenced by Edinburgh Genomics to determine whether they had the chocolate mutation. The 25 yellow dogs were chosen as follows: ten dogs with the highest minimum weight after one year of age (five male, five female); ten dogs with the lowest maximum weight after one year of age (five male, five female) and five dogs with the highest minimum ratio of weight to height^2 after one year of age (3 male, 2 female). DNA extraction, sequencing and analysis of the sequence was done by experimenters who were blinded to the relative weights of the dogs.

4.2.7 Neutering Information Cleaning & Analysis

Each time an owner answered the Dogslife questionnaire, they were asked whether their dog had been neutered and if so, when it was neutered. Once neutering information was given, they were not asked again. The neutering dates needed to be checked for obvious inaccuracies such as being neutered before the dog was born or being neutered in the future (relative to the time of answering the questionnaire).

The proportion of the cohort being neutered and their age of neutering was assessed and examined in the context of the dogs' sex.

4.2.8 Diet Cleaning & Analysis

Owners were asked about the type and quantity of food given to their dogs (as described in Section 3.3.2.2; questions available in Appendix 2, Section A2.5). They were asked for a numeric quantity value and given the opportunity to choose the relevant unit; g or ounces (oz). Amounts entered in ozs were multiplied by 28.35 and recorded in the database in grams. Quantities were initially examined for zero and non-numeric answers that were amended or treated as missing values (NAs). All quantities over 2000g were investigated in the context of other answers for that dog and answers such as 450400 were converted to 425g. Some answers appeared to be subject to a an error whereby the owner had chosen the incorrect units. For example, one measure of 8504.857 was likely 300g but entered as 300oz. These entries were divided by 28.35.

The entries whereby owners recorded their dog's diet type as 'other' required cleaning. These were visually examined and re-coded into those that included raw meat or eggs and those that did not.

Multiple rows of data might refer to one answer; for example a mixed diet would have a row relating to wet food and a row relating to dried food. These multiple rows had to be amalgamated into one row for future analyses. This was done on the basis of the pet ID and the time of reporting. Amalgamating the quantities was complicated by needing to add wet food weights to dry food weights. Future comparisons would be limited by the higher relative water content in the non-dried foods.

Once the details of mixed diets were amalgamated into a single row, the dietary data were then subjected to the technique that removed answers if the owner had entered new data within five days of their previous answer.

As discussed in Section 3.3.2.2, validation work indicated that diet data captured by the Dogslife questionnaire was not a perfect reflection of the true dietary intake of the dogs. Analysis of the collected data was therefore limited. The reports of types of diet were analysed and the quantities were examined graphically. The brands of commercially available food were not examined in detail. Feeding guides for four of the top ten most frequently reported brands of food were examined to determine how much the manufacturers advised a dog of 20-30kg should eat. On average, they advised 3.84 times as much wet food by weight as dry food. This factor was used to create an approximate dry food weight for the wet, mixed and other food types.

4.2.9 Exercise Cleaning & Analysis

The exercise data theoretically should have required minimal cleaning because all owner entries were chosen from drop-down menus that did not have 'other' options (questions available in Appendix 2, Section A2.4). There were no free-text entries so all that was required was for duplicate entries to be deleted and for entries where owners had corrected their previous answer to be given precedence. This latter was done in a similar fashion to Step 1 of both the weight and height cleaning, by deleting entries that had subsequent entries within five days of their date of reporting.

The exercise data were available as answers to individual questions. How much time, per day, did the dog spend on specific activities such as 'obedience training' on a weekday and on a day during the weekend. A completed questionnaire would comprise six lines of time categories, weekday and weekend amounts for each of six different types of activity. These amounts of time might then either be treated as ordered categorical variables, or given continuous values at random from appropriate uniform continuous distributions. Continuous times were created for each entry during the cleaning process both at random and as simple mid-points. For the final category of 'over 2 hours' a maximum of four hours was assumed making the mid-point three hours.

In order to reduce the number of variables associated with the exercise questionnaire, weekday and weekend measures were aggregated to produce a weighted average of the time spent on each activity. Weighting was done by multiplying the week day levels by $\frac{5}{7}$ and the weekend levels by $\frac{2}{7}$, then summing the two.

There were additional inexplicable oddities in the data. In theory, each time an owner answered the questionnaire, a unique Data Entry identifier (DEID) would be generated which would link together the answers to different sections of the questionnaire such as exercise and dog weight. Each DEID should only be associated with one set of questionnaire answers so exercise information given on date X would link to dog weight information also given on date X. Odd duplicates whereby sequential DEIDs were associated with duplicated information were recognised and dealt with early in the data cleaning process. It appeared however that some DEIDs were associated with multiple, different exercise entries. This was not a simple matter of an owner changing their mind during a data entry because each entry was dated and entries with the same DEID were separated by as much as five months. Once discovered, it was determined that this type of error was also present in other data types. For analysis purposes, each instance needed to be identified because data in these cases could not be linked using the DEID. If a data entry was associated with two different time periods, A and B, linking by the DEID would link A with A, A with B and B with B. Instead these entries had to be disregarded or linked by hand (depending on the practicality of hand-coding).

Reported exercise was considered in three different ways; as categorical variables, randomly generated continuous variables and continuous variables taken from the mid-point of each time category. The categorical reports given by owners were described and weekday and weekend activity levels were compared. The weekday and weekend times were aggregated and time categories were converted to randomly generated continuous time variables for visualisation purposes. Changes in exercise with age examined. Finally an aggregated weekday and weekend total time spent on all activity was generated by using the mid-point of the time categories. These total times spent were investigated with reference to other reported information such as dog location, dog purpose etcetera.

Each time the owners were asked for exercise details, they were also asked about exercise restrictions. Correlations between exercise restrictions categories and dog age were examined then the time spent exercising was assessed with reference to stated restrictions. Finally, a linear mixed models approach was taken to investigate total daily time spent exercising whilst taking into account the repeated measures. Both owner and pet IDs were included as random terms and a first order autocorrelation structure was fitted.

4.2.10 Sleeping Locations Cleaning & Analysis

Owners were asked to describe where their dog slept at night (question available in Appendix 2, Section A2.3) and initial cleaning involved deleting duplicates and removing entries that owners had changed themselves. At the inception of Dogslife, owners were given the following options: alone in a room in a house; in a room shared with a member of the family; outside and other. These options were amended in November 2013 as described in Section 3.3.2.1 with the aim of reducing the number of free-text answers. During the cleaning process, the proportion of free-text answers were examined with reference to that change. All free-text answers were then grouped according to the following categories:

- Alone
- With person - (In a room shared with a person and In a room shared with a pet and a person)
- With other pet only
- Outside

Sleeping locations were tabulated and examined in the context of the age of the dog and where they lived.

4.3 Results

4.3.1 Data Entries & Data Entry Windows

As mentioned in Section 2.3.2, 4,307 dogs were registered with Dogslife. The numbers of times owners answered the questionnaire for their dogs was captured by the number of data entries recorded in the database. Each time an owner began the questionnaire, theoretically, a unique DEID would be generated associated with the time and date when they began, which page of the questionnaire they reached (for example: 'page 1', 'page 2'... 'complete'), and the time and date they stopped entering information. By design, if an owner stopped entering data before the end of the questionnaire, the website would take them directly back to the page where they left off should they return within the same data entry window. The data entry windows were set differently according to the age of the dog. For dogs under one year, owners were asked to return every thirty days and would be considered to have answered within that window if they answered the questionnaire from 30 to 51 days. The next window would then be set relative to the timing of that data entry. For example if the questionnaire was answered at 40 days of age, the next data entry would be due at 70 days. For dogs over one year, the windows did not depend on the timing of the previous entry but instead spanned a period of 21 days before, to 35 days after, specific ages. These ages were at three month intervals after one year of age ie. 15 months, 18 months etcetera. Theoretically, a count of 'complete' DEIDs would give the number of completed questionnaires but there was an intermittent error such that an owner might complete the questionnaire but still be recorded as reaching one of the previous pages. This error meant that data entries were instead considered complete if there was an answer in the database to the final compulsory question in the questionnaire. The small number of data entries that were associated with multiple questionnaire answers (as described in Section 4.2.9) were disregarded in this calculation. There were 20,256 uniquely identified data entries within the database associated with 3,709 dogs. This implies that owners of 598 dogs registered with Dogslife but never answered the questionnaire. A further 460 dogs had one incomplete data entry, leaving data relating to 3,249 dogs for analysis. Only 17,896 entries of the 20,256 unique data entries fit into separate data entry windows and the distribution of these data entries per dog is shown in Figure 4.3.

4.3.2 Time-At-Risk

Each dog with one or more complete data entries contributed to the overall Dogslife risk set. The overall time at risk comprised 3,098 dog years and is shown, broken down by age, in Figure 4.4. The evidence of the impact of recall decay on illness and vaccination reporting in Section 3.6.4 and the long return intervals described in Figure 2.6 indicated that the full 3,098 dog years should not be considered in isolation because owners were likely to forget signs of potential illness as time passed leading to an underestimation of incidence rates. Instead, windows of risk for each dog in the period prior to a data entry should be used on the premise that owners might better remember a a short period rather than a period of up to three years. The impact of different time windows on the cohort

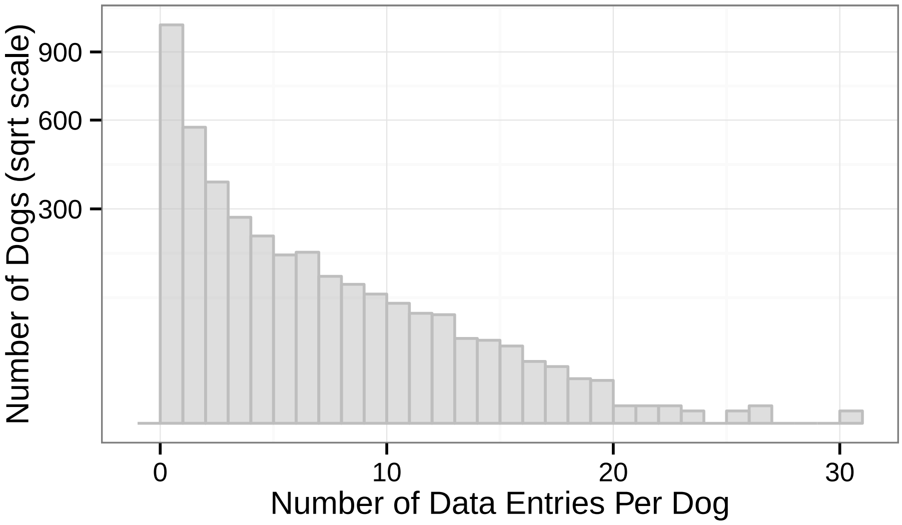


Figure 4.3: Data Entries (Questionnaires Answered) Per Dog. Number of dogs reported on a square root (sqrt) scale.

time at risk is shown in Figure 4.5.

The median time between veterinary visit and subsequent Dogslife data entry for illnesses that went unreported during validation was 40 days and this was the interval used to adjust the cohort time at risk. The adjusted time at risk was 1,733 dog years and a comparison of the time at risk (adjusted and unadjusted) is shown in Figure 4.6.

4.3.3 Cohort Age Structure

The seasonality of pup births is demonstrated in Figure 4.7 which shows the months of birth of the cohort. The distribution of dog ages at registration is shown in Figure 4.8. The target age for recruitment was under six months and it is noticeable that dogs were still being registered when well over this age.

4.3.4 Relatedness

Of the 4,307 Dogslife dogs, 14 were Guide Dogs about which Dogslife do not have information regarding parentage. Nineteen of the 4,293 remaining dogs for which pedigree information should have been available did not have a reported sire KCID in the UKKC records. All nineteen dogs had sire names and they were

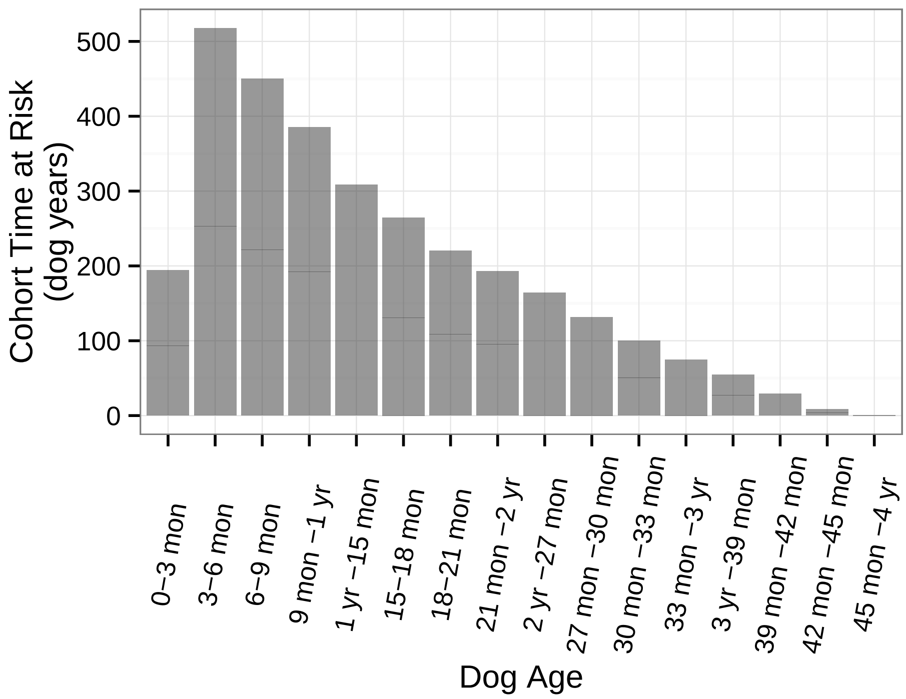


Figure 4.4: Cohort time at risk

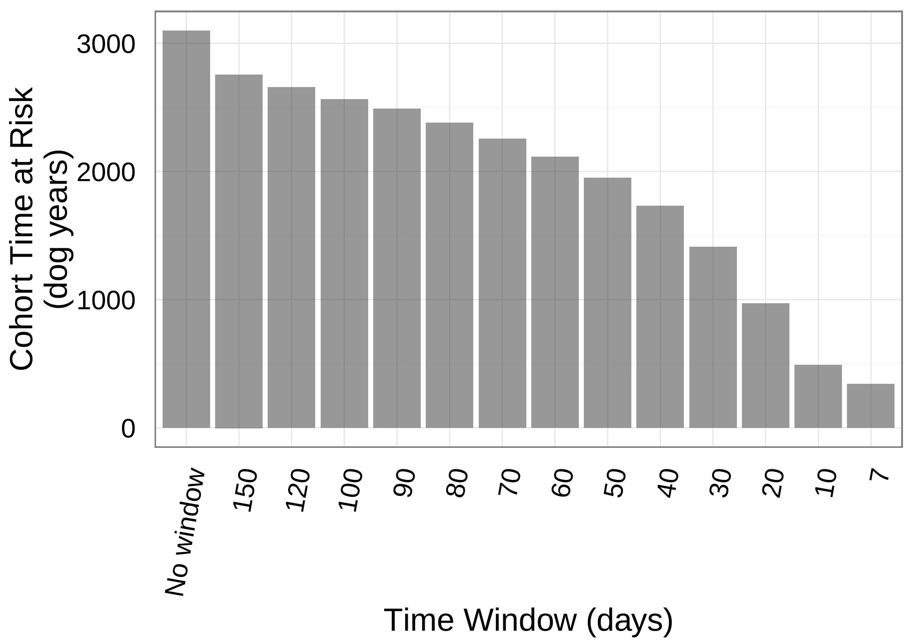


Figure 4.5: Impact of different time windows on overall cohort time at risk

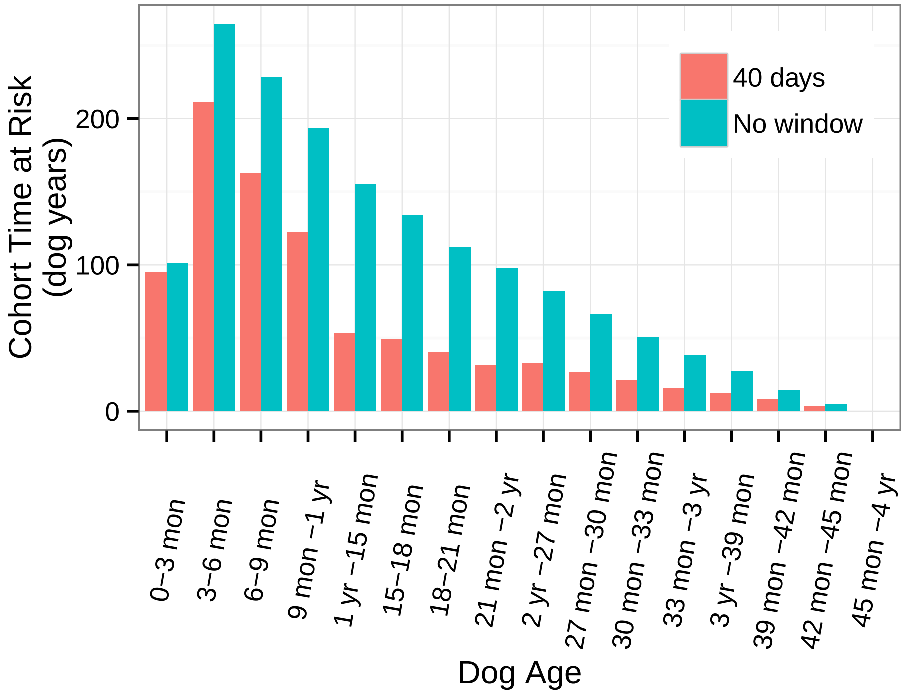


Figure 4.6: Time at risk adjusted using window of 40 days

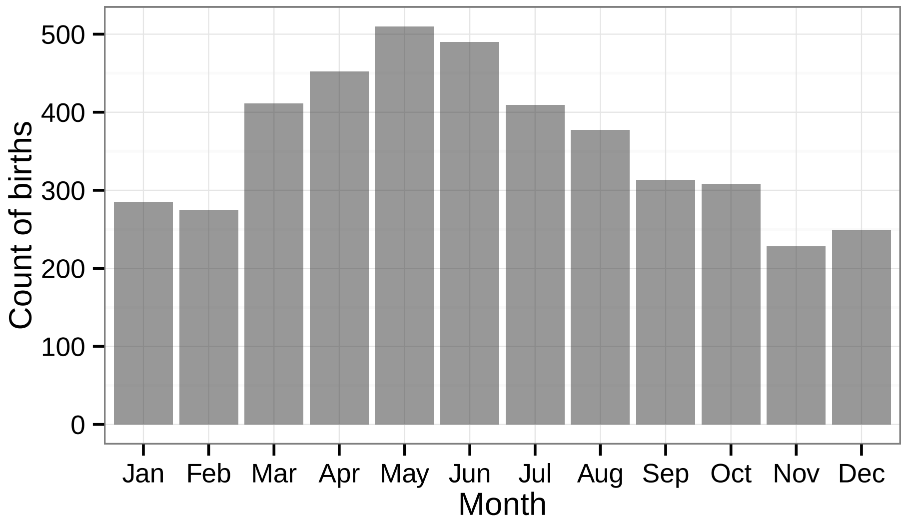


Figure 4.7: Cohort dates of birth split by months

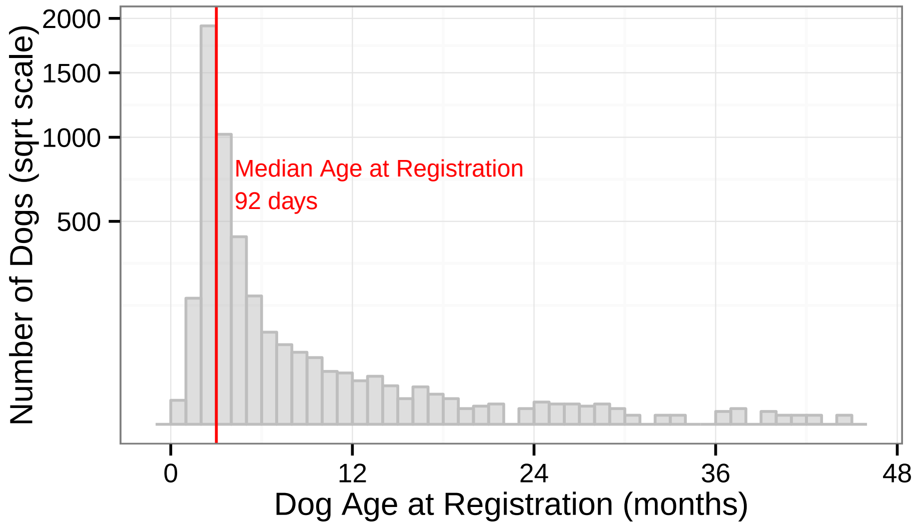


Figure 4.8: Dog Age At Registration

associated with just six sires. These sires came from USA (one sire contributing 11 Dogslife dogs), Ireland (one sire contributing two Dogslife dogs and another sire contributing one Dogslife dog), Finland (one sire contributing three Dogslife dogs) and France (one sire contributing one Dogslife dog). The final sire contributed one Dogslife dog and its origin could not be determined.

The degree of relatedness of the Dogslife cohort compared with 1,000 randomly selected groups of 4,293 dogs chosen from all eligible dogs is compared in Figures 4.9 and 4.10.

4.3.5 Dog Purpose

Owners were asked for the main purpose of their dog when they first began to fill in the questionnaire rather than at registration (as described in Section 3.3.1.4; question available in Appendix 2, Section A2.2). There were 35 different 'other' types of main dog purpose and after cleaning, the breakdown of different purposes was as follows: households pets (68%; 2,941/4,307), working dogs (5.8%; 253/4,307), assistance dogs (0.77%; 33/4,307), multi-purpose (0.46%; 20/4,307), show dogs (0.23%; 10/4,307), breeding dogs (0.046%; 2/4,307), other (0.56%; 24/4,307) and not reported (24%; 1,024). Show, breeding and multi-purpose dogs all appeared more than once but the majority of Dogslife dogs were primarily household pets. Many registered owners did not make a complete data entry

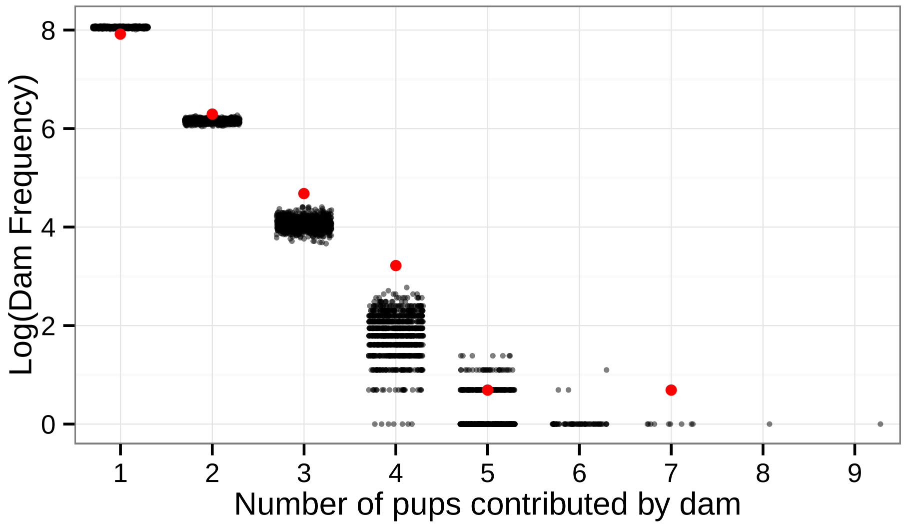


Figure 4.9: Comparison of number of pups contributed by the dams of Dogslife dogs and the dams of randomly selected UKKC dogs. The Dogslife cohort is shown in red and the dams of 1,000 randomly selected cohorts of UKKC dogs are shown in black.

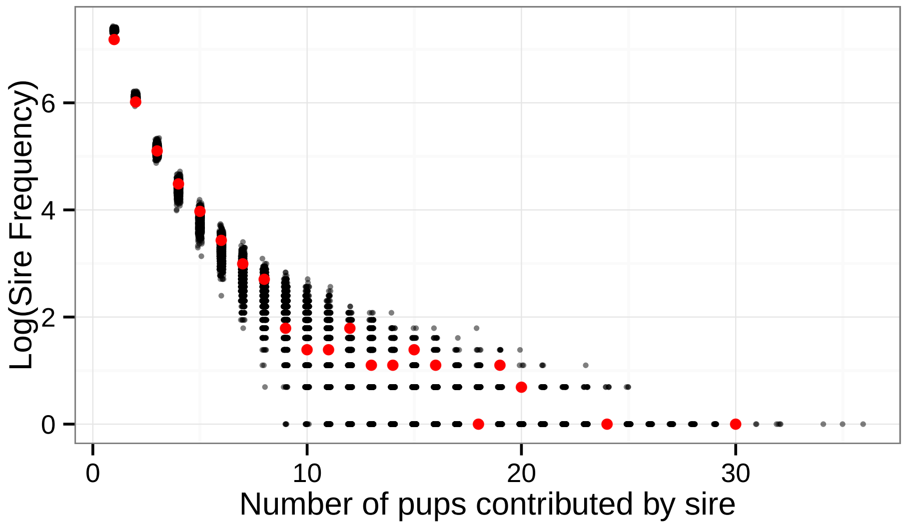


Figure 4.10: Comparison of number of pups contributed by the sires of Dogslife dogs and the sires of randomly selected UKKC dogs. The Dogslife cohort is shown in red and the sires of 1,000 randomly selected cohorts of UKKC dogs are shown in black.

so it is unsurprising that data are missing regarding the purpose of over 1,000 dogs. There were multiple associations between reported dog purpose and household type (Table 4.1). Working dogs were under-represented in family households and over-represented in households comprised of more than one adult. Assistance dogs were more likely to be found in retired households.

4.3.6 Dog Coat Colour

There were initially 23 different coat colours reported to Dogslife. The majority of these were variations on 'fox red' but also included 'champagne' and 'white chocolate'. Dogslife had access to the UKKC records of coat colours for the cohort and these reported colours guided the cleaning process. For example, the 'white chocolate' dog was 'yellow' according to the UKKC so it was considered 'yellow' in the cleaned coat colour data. The frequencies of cleaned colours are shown in Table 4.2. The UKKC do not recognise Hailstone or Fox Red as legitimate coat colours and there were further discrepancies between Dogslife reported colour and UKKC registered colours. These might have been data entry errors or a mix-up between puppies in litters. There were also isolated oddities such as the dog shown in Figure 4.11 whose owner reported it to Dogslife as being 'black and tan' but who was registered with the UKKC as simply 'black'. For ongoing analysis purposes, where owners had reported 'other' or they failed to give a colour ('not given'), then the UKKC colour was used.

It is not strictly legitimate to directly compare the proportions of each coat colour registered with Dogslife with those eligible to join the project because the UKKC offered fewer coat colours options. Nevertheless, when all those that were not recognised by the UKKC were grouped together, it appears that the rarer coat colours such as chocolate and fox red were over-represented in the Dogslife cohort (Figure 4.12).

Table 4.1: Associations Between Dog Purpose & Household Type

	Family (%)	More than one adult (%)	Retired (%)	Single adult (%)	Not given (%)
Household pet*	1288	1231	205	153	64*
	(43.8)	(41.9)	(7.0)	(5.2)	(2.2)
Working dog*	84*	132*	21	9	7
	(33.2)	(52.2)	(8.3)	(3.6)	(2.8)
Assistance dog[F]	8	11	10[F]	3	1
	(24.2)	(33.3)	(30.3)	(9.1)	(3.0)
Multi-purpose[F]	7	9	2	2	0
	(35)	(45)	(10)	(10)	(0)
Show dog	3	4	0	3	0
	(30)	(40)	(0)	(30)	(0)
Breeding dog	1	1	0	0	0
	(50)	(50)	(0)	(0)	(0)
Other[F]	8	8	2	4	2
	(33.3)	(33.3)	(8.3)	(16.7)	(8.3)
Not given*	515*	350*	47*	61	51*
	(50.3)	(34.1)	(4.6)	(6)	(5)
Total	**1914**	**1746**	**287**	**235**	**125**
	(44.4)	(40.5)	(6.7)	(5.5)	(2.9)

* χ^2 test with Bonferroni correction indicates association, $P < 0.00\dot{3}$
* Due to low numbers in many categories, only household pet, working dog and household not given categories were assessed for associations.
[F] Fisher's exact with Bonferroni correction indicates association, $P < 0.00\dot{3}$
[F] Due to very low numbers, show and breeding dog categories were not considered.
Blue text indicates negative association and red indicates positive association.

Table 4.2: Dog Coat Colours

Dogslife Coat Colour	Number *(%)*	UKKC Registered Colour
Black	2121	2113 black
	(49.2)	5 yellow
		1 chocolate
		2 guide dogs[a]
Yellow	1167	1162 yellow
	(27.1)	4 black
		1 liver
Chocolate	898	873 chocolate
	(20.8)	25 liver
Fox Red	96	95 yellow
	(2.2)	1 not recognised by KC
Hailstone	1	1 black
Black & Tan	1	1 black
Other	14	0 black
	(0.3)	13 yellow
		1 chocolate
Not Given[b]	9	5 black
		3 yellow
		1 chocolate

[a] GDBA have their own pedigree information and their dogs are not registered with the UKKC.
[b] A free-text box is generated when an owner chooses 'other'. This box may then be left blank.

Figure 4.11: Black & Tan Labrador

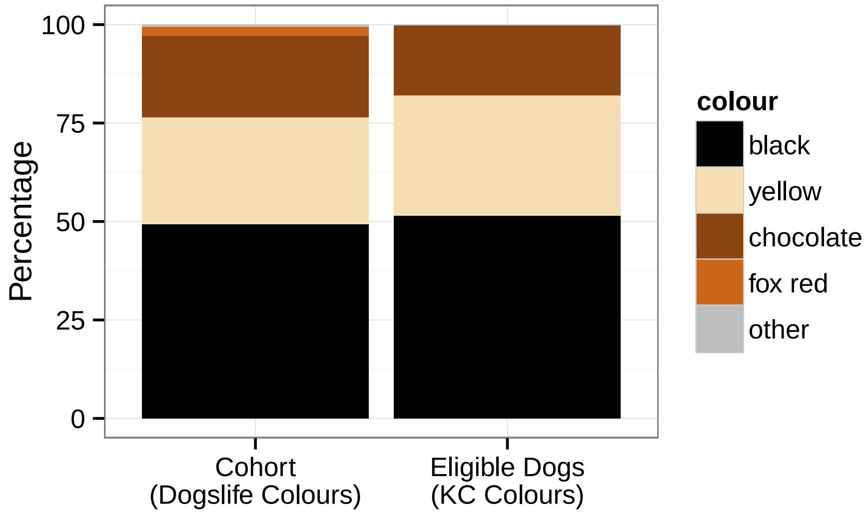


Figure 4.12: Proportion of each coat colour in the Dogslife cohort compared with coat colours in UKKC registered LR

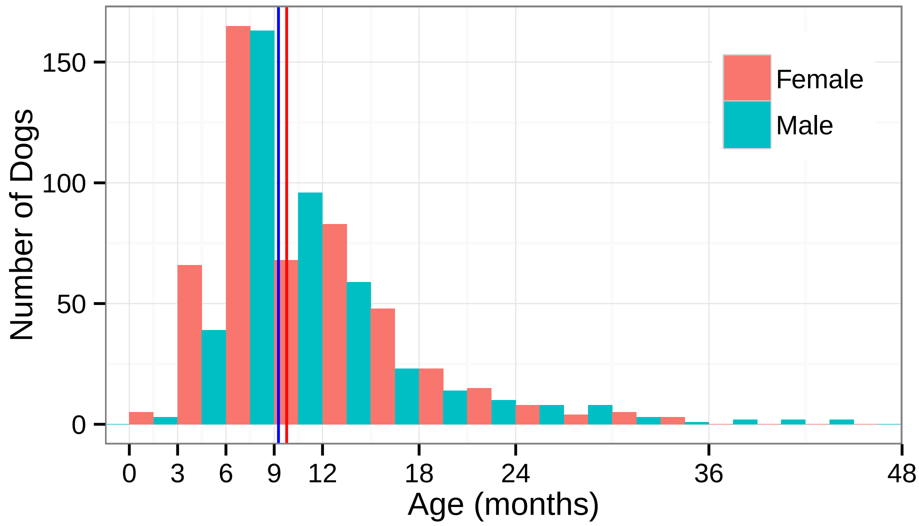


Figure 4.13: Age of neutering split by sex
Median ages are shown using vertical lines for in blue for males and red for females

4.3.7 Neutering

Neutering data were collected for 2,969 dogs and of these, 924 were reported as being neutered. Twenty three neutering dates were in the future relative to the date of reporting. Seven appeared to have the wrong year and these were amended. Fifteen owners had given the month of data entry rather than the month of neutering but in the absence of further information, these dates were not amended. One owner had given a date that made no sense and attempts were made to contact this owner for further information. Three owners had given dates that were before their dog was born. These were all assumed to be out by a year and were amended.

For males, the neutering rate was 28.1% (95% CI: 25.8 - 30.4%) increasing to 34.5% (95% CI: 32.0 - 37.0%) for females. The distribution of age of neutering is shown in Figure 4.13. The median neutering age is 282 days for males and 297 days for females (ranges 35 - 1,349 and 33 - 1,077 days respectively).

There were 2,045 dogs who were not reported to have been neutered but there is strong evidence of loss to follow up in these dogs. When owners reported that their dog had been neutered, they were no longer asked about neutering. Despite this, the 924 owners who reported neutering answered the neutering question

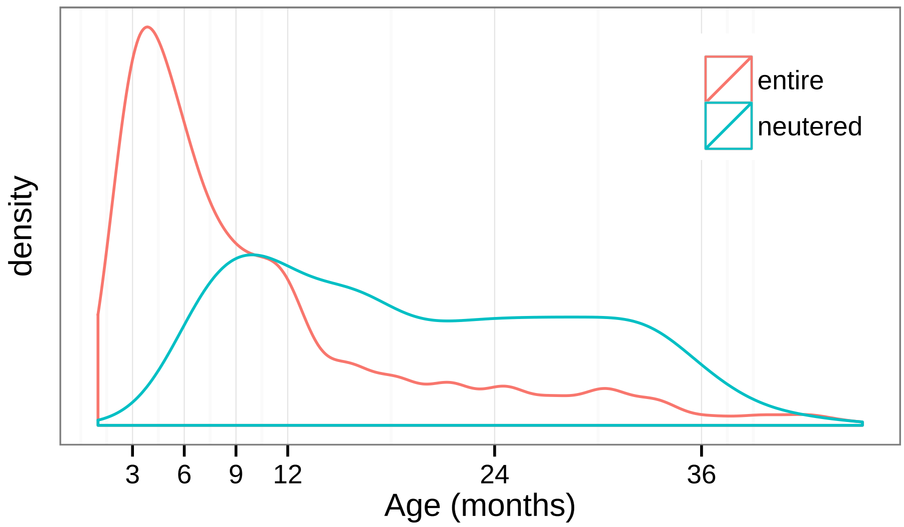


Figure 4.14: The red density plot shows the age at last data entry for un-neutered dogs. The blue density plot shows the age at neutering for neutered dogs. Rather than using histograms which would be dominated by the more numerous un-neutered dogs, density plots have been used to facilitate comparison between the two groups.

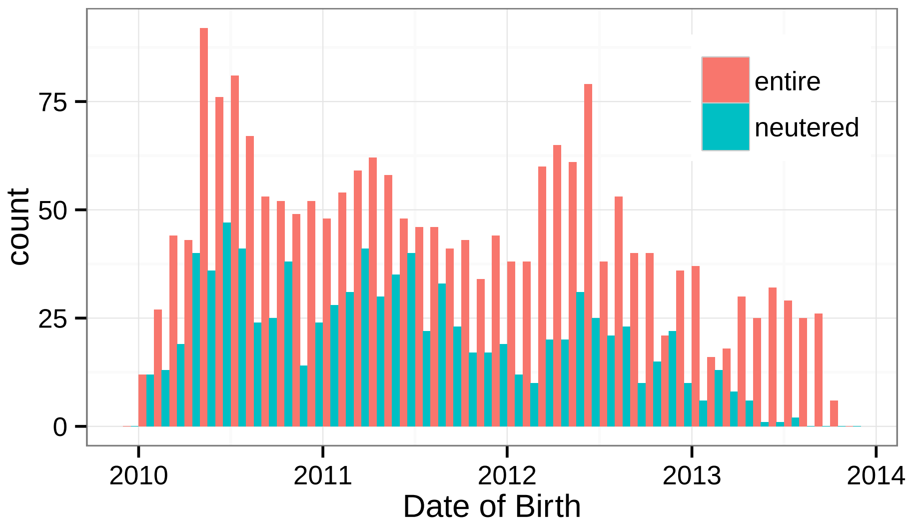


Figure 4.15: Dates of birth of the neutered and entire dogs

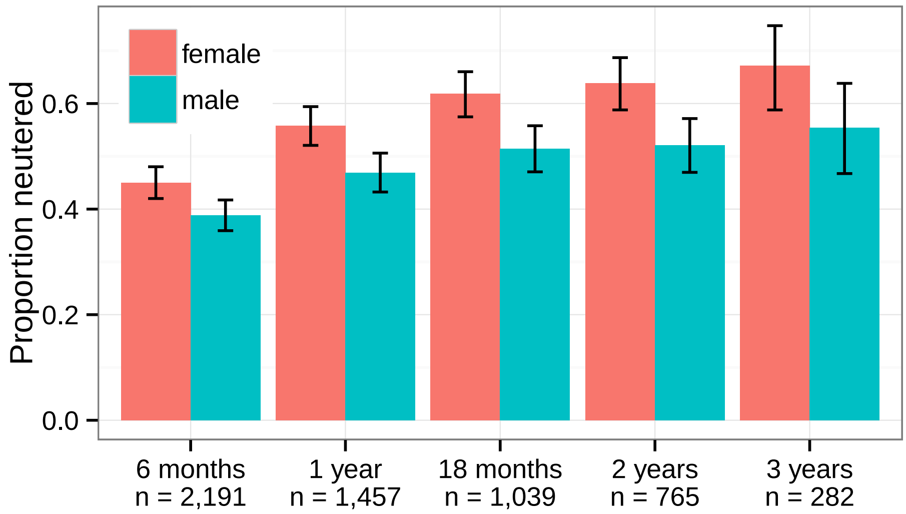


Figure 4.16: Cumulative neutering rates (with 95% CI) for cohort members that have associated data entries after each given age

over six times on average, whereas those that did not report neutering answered it just three times. Figure 4.14 compares the age of neutering for the neutered dogs with the age of the most recent contribution to Dogslife for the dogs with no reported neutering. The peak of this latter group lies firmly to the left of the peak of neutering age. As Figure 4.15 demonstrates, the entire dogs included a group that were slightly younger than the neutered dogs (born from spring 2013 onwards) but this group cannot entirely account for the difference between ages in Figure 4.14. It is possible that many of the dogs that were not reported to have been neutered were neutered after their most recent Dogslife entry. The estimates of neutering rates given previously should therefore be regarded as underestimates.

If only dogs that had associated data entries after six months, one year, eighteen months, two years and three years were considered then the cumulative neutering rates increase markedly (Figure 4.16). What remains unclear is whether these higher neutering rates reflect the true rate in UK LRs or whether owners who stay part of the project are more likely to neuter their dogs.

4.3.8 Diet

The reported dietary information comprised over 18,500 rows of data. This included multiple rows for entries regarding mixed diets. Consolidation of these mixed diets into single rows left over 16,200 entries and after these were amended to account for owner corrections, approximately 15,200 unique entries remained. Some 753 rows related to 'other' type diets which could be split into 290 rows of diet that included raw meat or eggs and 463 rows of 'other'.

There were 2,222 unique brands of commercial food but this included multiple different spellings and a small number of mistakes whereby people had entered quantities in the wrong box.

Diet data were reported for 3,097 dogs, of which 806 dogs appeared only once. The frequencies of the different types are shown in Table 4.3. If the dogs that only appeared once are disregarded then 2,291 dogs remain and it is possible to assess whether their diets changed. The majority (1,642) did not have varying diet types, 1,503 of them eating a consistent diet of dried food. The reported dietary quantities are plotted in Figure 4.17 and the plot where all non-dried content was divided by 3.84 is Figure 4.18.

Table 4.3: Frequencies of different diet types

Type	Frequency
Dried	12,124
Mixed	2,005
Raw	291
Home prepared	171
Wet	165
Other	463

The majority of reports involved dogs being fed twice per day, in the morning and evening but at younger ages dogs were fed more often (Figure 4.19).

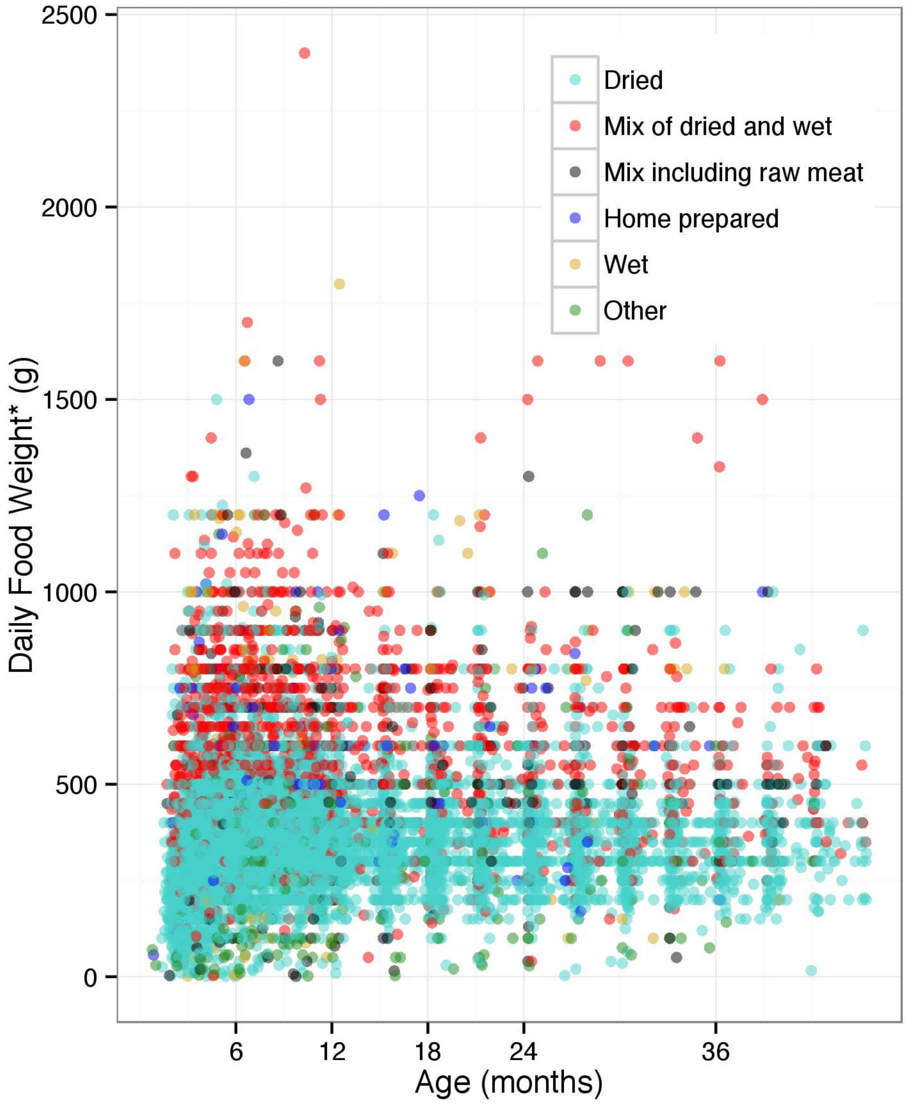


Figure 4.17: Food Weights By Age
* The wet, mixed, raw, home prepared and other foods all have a higher moisture content than dried alone.

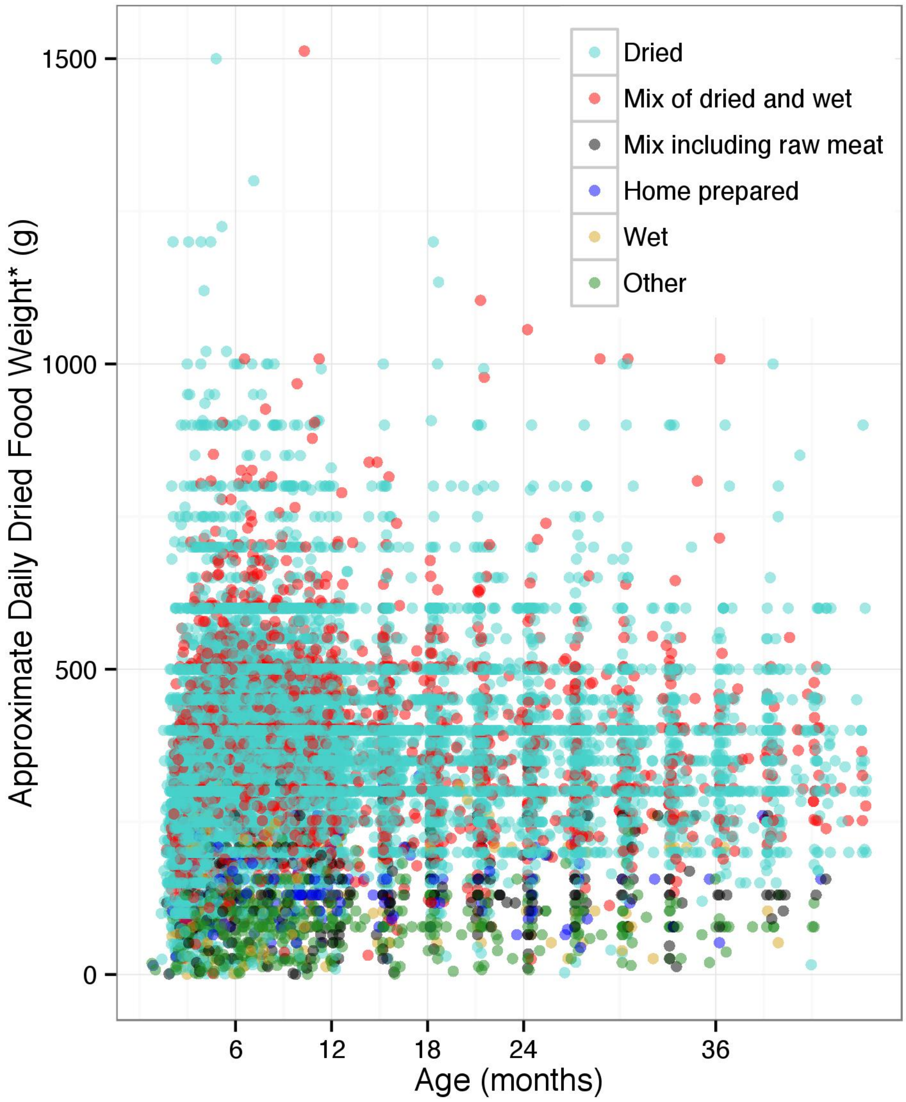


Figure 4.18: Food Weights By Age
* The wet, mixed, raw, home prepared and other foods all had their non-dried food content divided by 3.84.

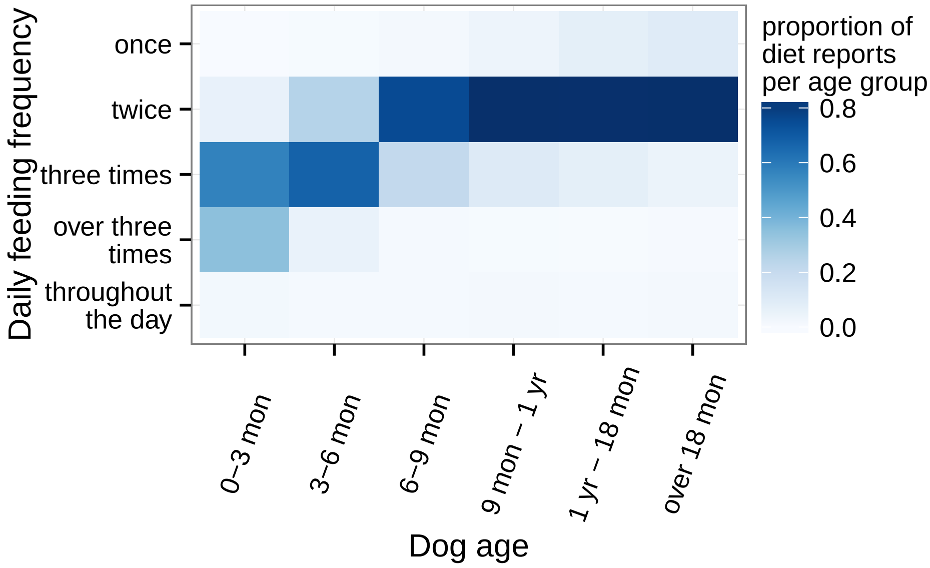


Figure 4.19: Proportion of diet reports for each age group that fall into each frequency group

4.3.9 Dog Heights

There were 15,764 height measurements relating to 3,202 dogs and they are shown, plotted against the ages of the dogs in Figure 4.20. Step 1 of the cleaning process reduced this number to 12,600 heights that are shown in Figure 4.21. Step 2 involved amending 10 heights and Step 3 removed a further 61 entries, leaving 12,539 measurements (Figure 4.22).

The Bayesian estimating procedure amended 470 heights; 455 were deemed 2.54 times too low and 15 were deemed too high. Figure 4.23 shows the merging of what had previously been two growth curves. Finally, Step 5 involved the amendment of heights of greater than or equal to 75 cm. Ten appeared to be the result of mis-typing and were amended; 37 were divided by 2.54 and 60 heights were deleted leaving 12,479. The final cleaner data are shown in Figure 4.24.

The model checking plots are shown in Appendix 4. Parameter pi[3] is the proportion of height measurements that were 2.54 times too high. This value was so small, its value was not definitively positive according to the model and it therefore could not be included in the Gelman Plots.

The dog heights are shown split according to sex in Figure 4.25 with the parameters of the modelled growth curves (Equations 4.3) detailed in Table 4.4.

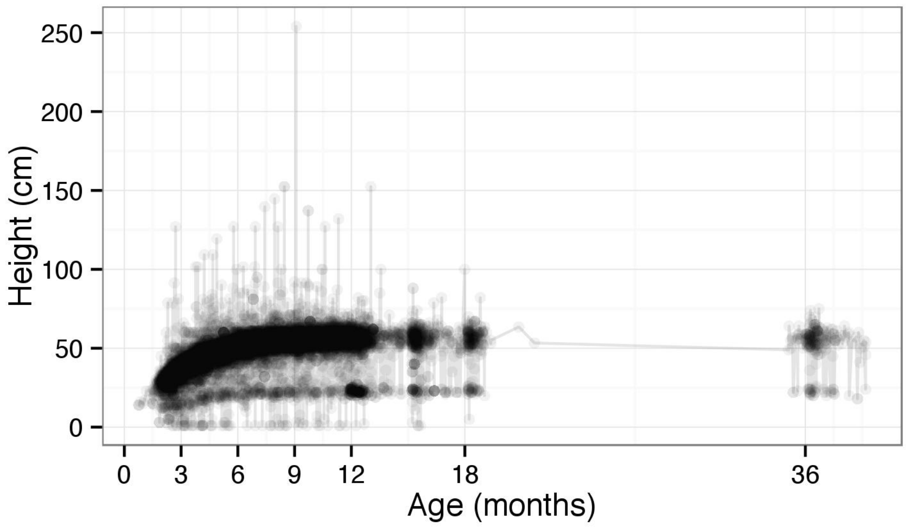


Figure 4.20: Heights By Dog Age (Uncleaned)

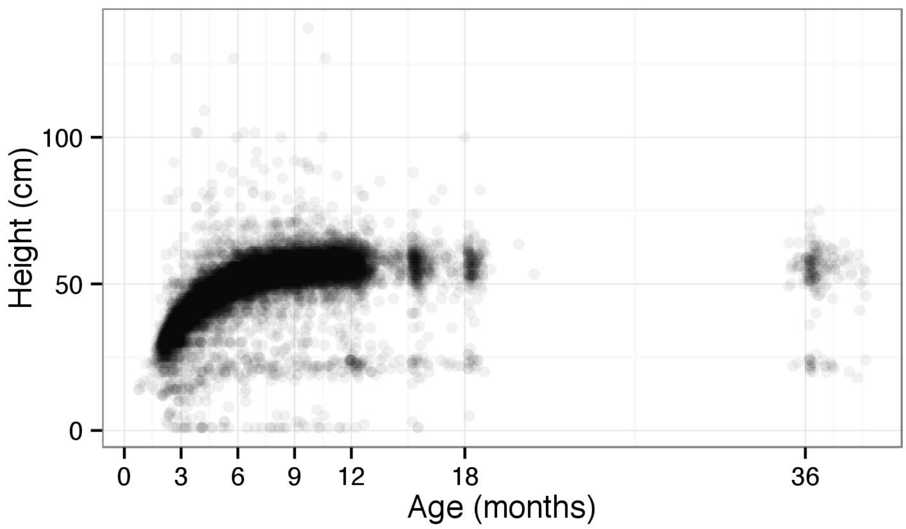


Figure 4.21: Heights By Dog Age (Cleaned Step 1)

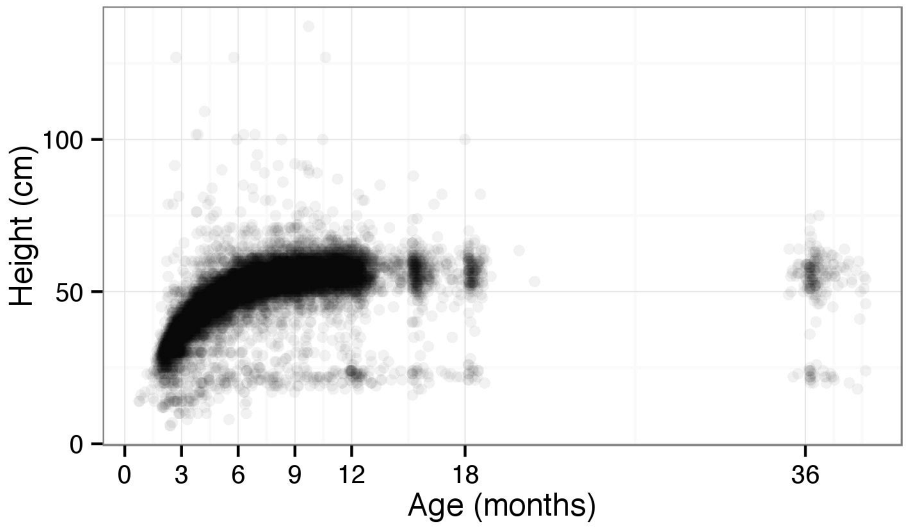


Figure 4.22: Heights By Dog Age (Cleaned Step 3)

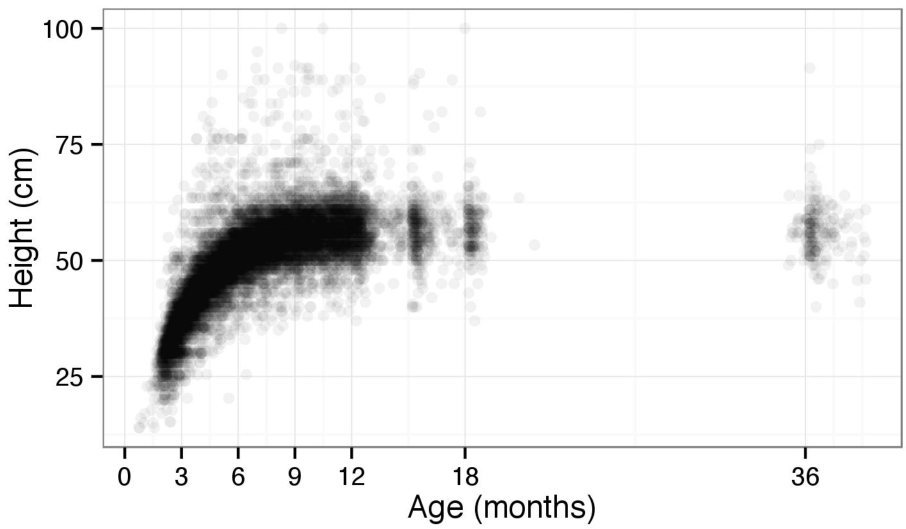


Figure 4.23: Heights By Dog Age (Cleaned Step 4)

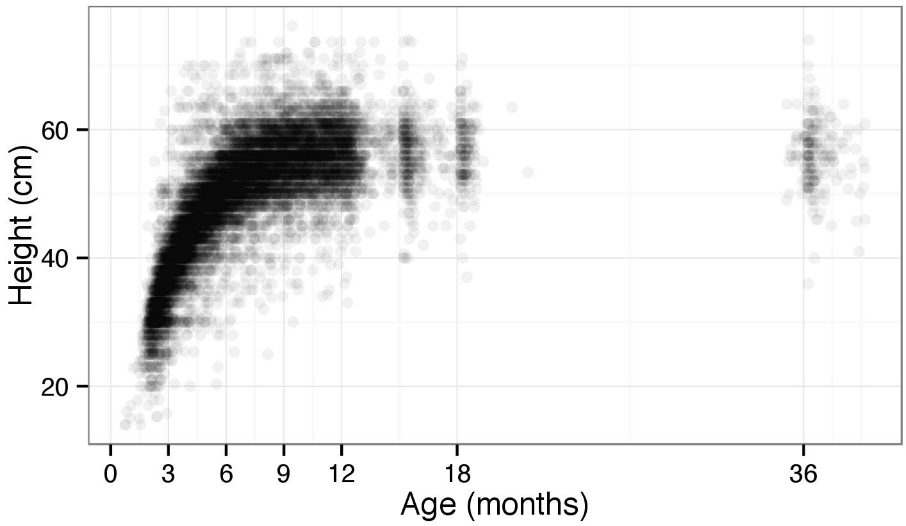


Figure 4.24: Heights By Dog Age (Final Cleaner Data)

Table 4.4: Dog height growth model parameters

Variable	Female (95% CI:)	Male (95% CI:)
A	55.1 (54.9 - 55.4) cm	59.0 (58.7 - 59.2) cm
B	0.0132 (0.0128 - 0.0137)	0.0126 (0.0122 - 0.0131)
C	7.03 (4.43 - 9.63) days	9.37 (6.77 - 11.9) days
τ	4.67 (4.59 - 4.76) cm	5.01 (4.92 - 5.10) cm

$$height = \mathcal{N}(\mu, \tau) \tag{4.3a}$$

$$\mu_i = A(1 - e^{(-B(age_i - C))}) \tag{4.3b}$$

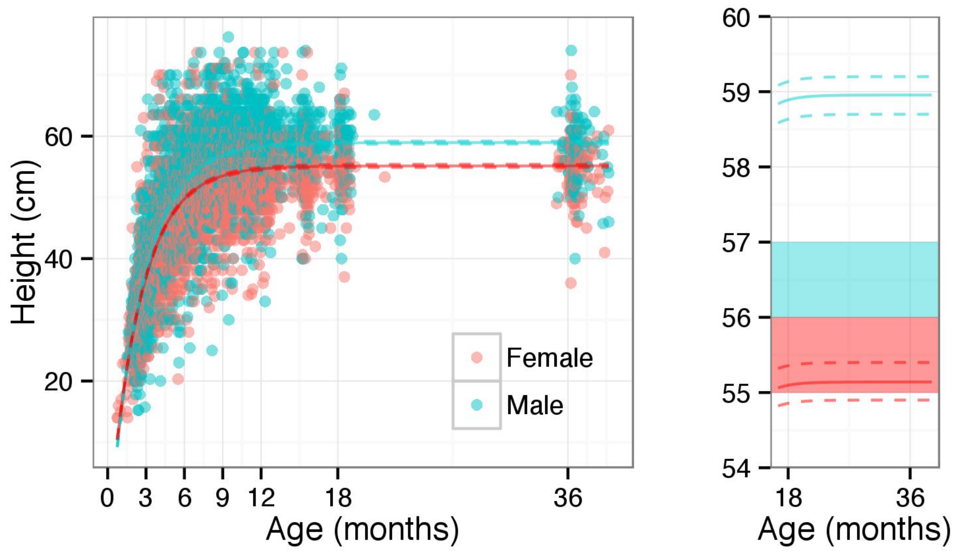


Figure 4.25: Modelled dog heights with dashed 95% credible intervals for parameter A. The solid bands of colour in the expanded right-hand figure show the UKKC breed standard for each sex.

4.3.10 Dog Weights

There were 19,654 weights reported for 3,196 dogs. These data are shown in Figure 4.26 plotted against dog age. It was immediately apparent that tonne-range weights were errors whereby owners likely made a measurement in grams but reported in kg. Step 1 of the cleaning process removed one of the tonne-range weights and reduced the overall number of weights to 16,065. Step 2 dealt with a further three measurements over 1000 kg and 16 measurements over 100 kg. Twelve of those over 100 kg were simply divided by 10 but measurements of 128 and 128.5 became 28 and 28.5 respectively. A measurement of 555 became 5.5 and a measurement of 124.74 could not be reconciled with the rest of that dog's record and was deleted. These partially cleaned data are again plotted against age in Figure 4.27.

Step 3 of the cleaning process identified 43 weights below 17 kg when the dog was aged over 300 days. One dog's entire record (three weights) was deleted and another five dogs had single 'nonsense' entries deleted. Three easily-identifiable typographical errors were corrected. The remaining 27 dogs were examined as potentially subject to lb-kg units errors. Eight were assumed to be light-weight dogs and 19 dogs had low weights multiplied by the conversion factor as described

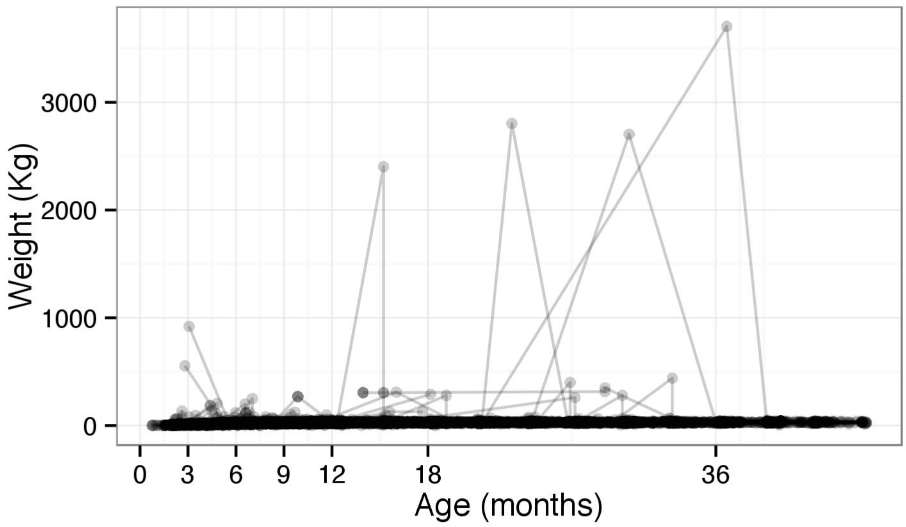


Figure 4.26: Weights By Dog Age (Unclean)

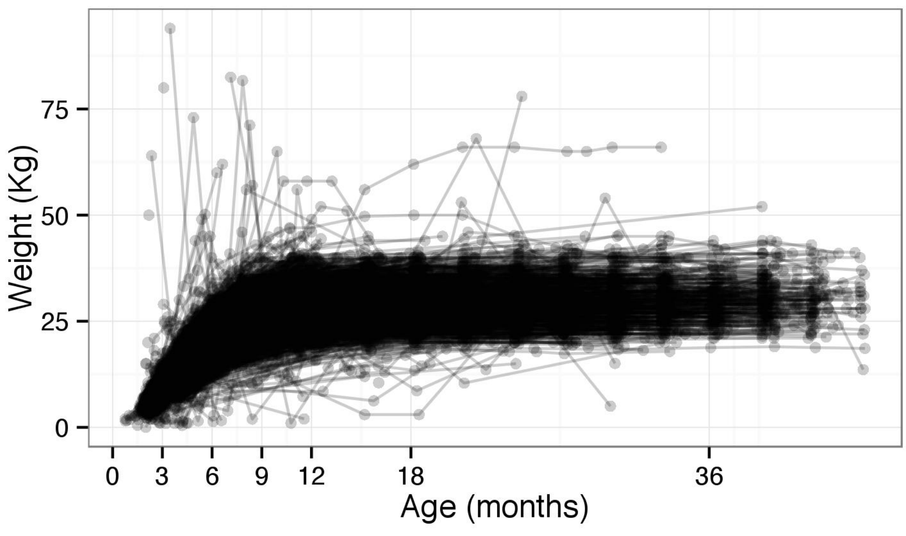


Figure 4.27: Weights By Dog Age (Cleaned Step 2)

in Section 4.2.5. A further 11 weights were below 12 kg when the dog was aged over 200 days. Six dogs had weights multiplied by the conversion factor. One dog had a weight deleted and three dogs were assumed to be light-weight. Finally, 19 weights were below 5 kg when the dog was aged over 100 days. Six weights and one whole record were deleted, six weights were amended due to assumed typographical errors and five weights were multiplied by the conversion factor. Two measurements were consistent with the rest of the weight record for those respective dogs and were left. The latest stage of cleaning is shown in Figure 4.28.

The final step of cleaning the dog weights involved examination of relatively high weights (Step 4). Some 31 weights over 50 kg were identified associated with 19 dogs. Three weights were deleted (one was the dog's height). Five typographical errors were amended and weights were divided by the conversion factor for seven dogs. The weights for three of the dogs had been steadily increasing with time and these were not amended. The tidy weights (16,043) are shown in Figure 4.29 and these were used in all future analyses.

Cleaned dog weights at all ages are shown split according to sex in Figure 4.30 and according to coat colour in Figure 4.31. The results of the univariable analysis seeking associations between unchanging factors and the weights of dogs over one year are shown in Table 4.5.

Table 4.5: Univariable analyses of associations between non-changing factors and the weights of dogs (kg) aged over one year

Variable	Coefficient	95% CI: lower	upper	P value
Mature Height (cm)	0.412	0.36	0.47	<0.0001
Sex				
female	-	-	-	-
male	4.79	4.4	5.2	<0.0001
Coat colour				
black	-	-	-	-
chocolate	1.64	0.99	2.3	<0.0001
yellow	0.119	−0.48	0.72	0.70
fox red	−0.888	−2.6	0.84	0.32

Continued on next page

Table 4.5: Univariable associations (continued)

Variable	Coefficient	95% CI: lower	upper	P value
Smoking status				
no	-	-	-	-
yes	1.55	0.82	2.3	<0.0001
Dog Purpose†				
pet	-	-	-	-
working	−2.94	−3.9	−2.0	<0.0001
other	1.66	−0.20	3.5	0.087
Household type				
family	-	-	-	-
more than one adult	0.726	0.14	1.3	0.016
retired	−0.579	−1.5	0.32	0.20
single adult	0.822	−0.30	1.9	0.15
not reported	−0.906	−2.8	1.0	0.36
Other Dog				
no	-	-	-	-
yes	-1.43	-0.90	-1.97	<0.0001
Country‡				
England	-	-	-	-
Wales	1.73	0.229	3.24	0.024
Scotland	−0.272	−0.975	0.432	0.45
Northern Ireland	−0.660	−2.74	1.42	0.53

†Assistance dogs were dropped from this analysis as they typically leave the Dogslife project at one year of age. All non-pets and non-working dogs were grouped as ‘other’.

‡There were too few dogs from Jersey, Guernsey and the Isle of Mann to be included here.

In light of the finding that chocolate dogs were heavier than their non-chocolate counterparts, DNA testing was undertaken by Ailsa Carlisle and Edinburgh Genomics to determine whether heavier yellow dogs might carry the chocolate mutation. Of the 25 dogs chosen, the DNA was only of sufficient quality to test 21 (10 light, 11 heavy) and of these, none were homozygous for any of the chocolate alleles. Three of the heavier dogs were heterozygous at exon 4 but all other dogs were homozygous black (Figure 4.1, genotype eeBB).

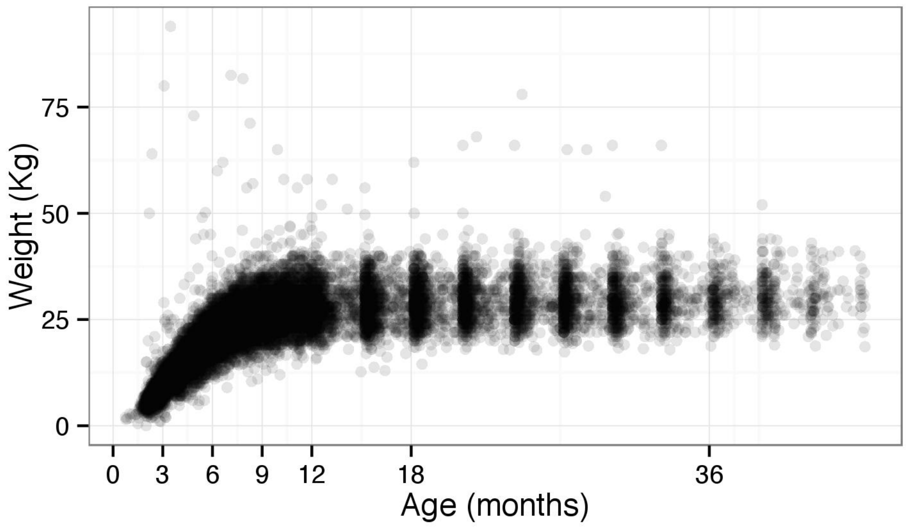


Figure 4.28: Weights By Dog Age (Cleaned Step 3)

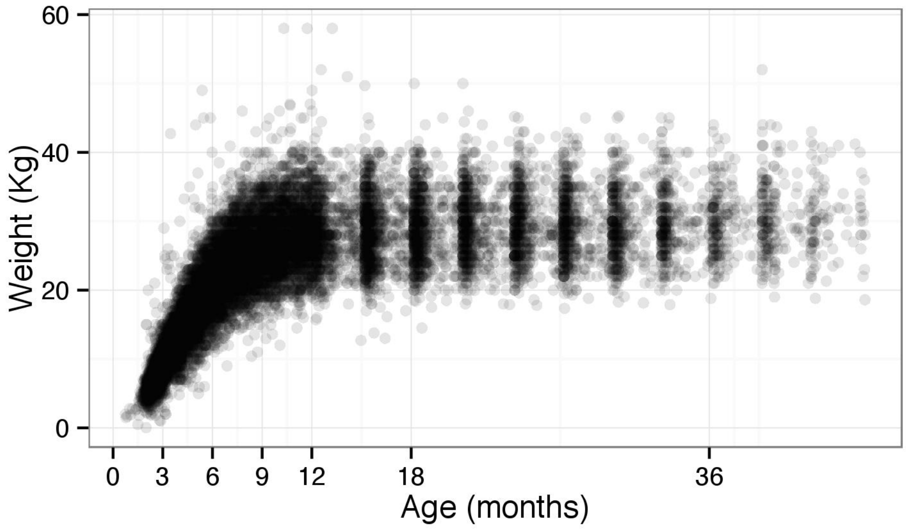


Figure 4.29: Weights By Dog Age (Cleaned Step 4)

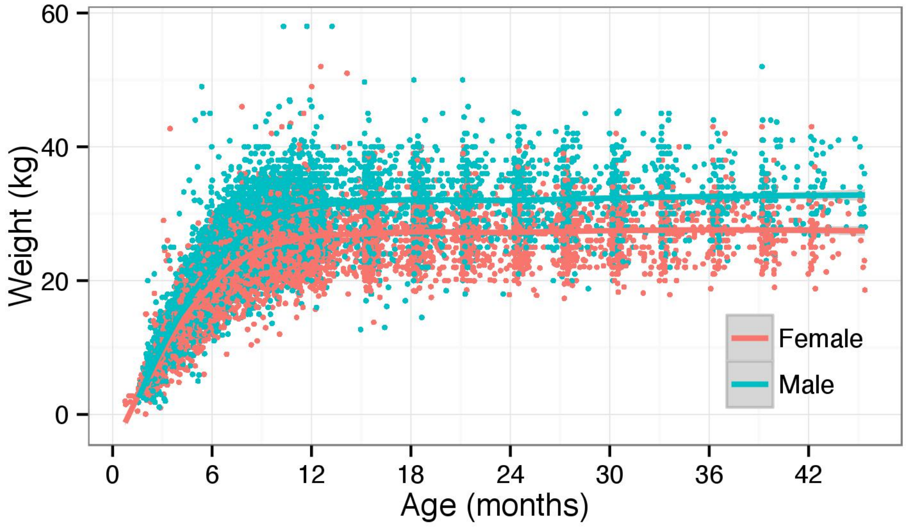


Figure 4.30: Dog Weights By Age & Sex
Coloured lines are locally weighted best fit smoothing lines for each sex

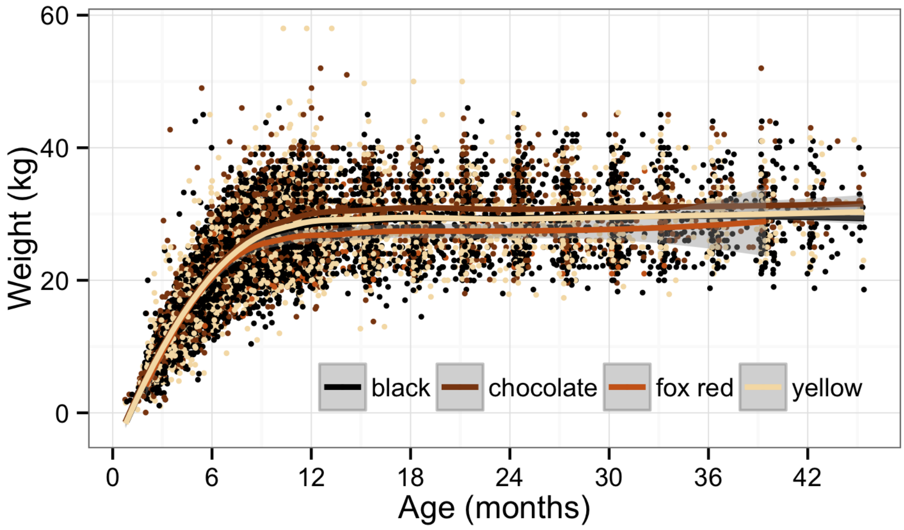


Figure 4.31: Dog Weights By Age & Coat Colour
Coloured lines are locally weighted best fit smoothing lines for each coat colour

Associations between dog weight and factors that changed with time such as dog age, exercise regime and diet were also considered in univariable analyses and the results are shown in Table 4.6.

Table 4.6: Univariable analyses of associations between changing factors and the weights of dogs (kg) aged over one year

Variable	Coefficient	95% CI: lower	upper	P value
Age (years)	0.798	0.70	0.90	<0.0001
Neuter Status†				
entire	-	-	-	-
neutered	8.82	8.53	9.11	<0.0001
Food Weight	1.81×10^{-4}	-4.53×10^{-4}	8.14×10^{-4}	0.58
Titbits				
no	-	-	-	-
yes	0.0866	−0.067	0.24	0.27
Food Type				
dried	-	-	-	-
other	0.439	0.097	0.780	0.012
wet	−0.144	−0.923	0.635	0.717
mixed	0.124	−0.150	0.397	0.375
Exercise restrictions				
unrestricted	-	-	-	-
recommended by breeder	−0.346	−0.607	−0.0858	9.23×10^{-3}
owner's health	0.745	0.283	1.21	1.61×10^{-3}
location	0.901	0.119	1.68	0.0240
owner's time	−0.138	−0.403	0.126	0.305
dog's health	−0.108	−0.490	0.274	0.580
Total daily exercise time (minutes)	-9.98×10^{-4}	-1.97×10^{-3}	-9.82×10^{-5}	0.0298

Continued on next page

Table 4.6: Univariable associations (continued)

Variable	Coefficient	95% CI: lower	upper	P value
Exercise daily time spent (minutes) ‡				
other	-2.59×10^{-3}	-4.44×10^{-3}	-7.25×10^{-4}	6.48×10^{-3}
lead walking	3.95×10^{-3}	1.08×10^{-3}	6.82×10^{-3}	7.02×10^{-3}
fetching, chasing & retrieving	-3.18×10^{-3}	-6.01×10^{-3}	-3.42×10^{-4}	0.0281
off lead	-1.44×10^{-3}	-3.38×10^{-3}	5.03×10^{-4}	0.147
obedience work	-3.49×10^{-3}	-8.23×10^{-3}	1.31×10^{-3}	0.154
lead running	-2.02×10^{-3}	-9.56×10^{-3}	5.52×10^{-3}	0.600

†As mentioned in Section 4.3.7, neuter status is very strongly correlated with age so the coefficients here should be treated with caution.

‡Using the midpoint of each time category as continuous data.

The results of a linear mixed model which combines these factors are shown in Table 4.7. The interaction terms were excluded because they did not improve the model. Fetching, chasing and retrieving and 'other' exercise were both associated with dog weight but the total time spent exercising was not. Working dogs, a group that typically spent more time exercising than pets, were more than 2 kg lighter than pets. The mean weight of a two-year-old Dogslife LR was 26.8 kg for females and 31.6 kg for males. Both measurements fit within the suggested weight range reported for adults of the breed of 25-34 kg (Alderton and Morgan [1993]).

Table 4.7: Results of linear mixed model of weights of dogs (kg) aged over one year

Variable	Coefficient	95% CI: lower	upper	P value
Intercept	18.40	16.80	19.90	<0.0001
Age (years)	0.890	0.762	1.02	<0.0001
Neuter Status				
Entire				
Neutered	-0.120	-0.366	0.126	0.339
Height^2 (cm)	2.24×10^{-3}	1.76×10^{-3}	2.72×10^{-3}	<0.0001
Coat Colour				
Black				
Chocolate	1.39	0.775	2.00	1.50×10^{-4}
Fox red	-0.843	-2.46	0.773	0.316
Yellow	0.189	-0.347	0.725	0.495
Sex				
Female				
Male	3.65	3.15	4.16	<0.0001
Dog Purpose				
Pet				
Working dog	-2.13	-3.01	-1.25	<0.0001
Other†	2.49	0.745	4.24	9.60×10^{-3}
Smoking Status				
Non-smoker				
Smoker	1.09	0.412	1.77	1.72×10^{-3}
Not reported	-1.40	-3.49	0.688	0.189
Other Pets				
No dog				
Dog	-0.477	-0.986	3.03×10^{-2}	0.07

Continued on next page

Table 4.7: Linear mixed weight model (continued)

Variable	Coefficient	95% CI: lower	upper	P value
Daily Time Exercising (hours)				
Fetching, chasing & retrieving	-0.218	-0.354	-8.2×10^{-2}	1.66×10^{-3}
Other	-8.64×10^{-2}	-0.181	8.16×10^{-3}	7.35×10^{-2}
Exercise Restrictions				
None				
Location	0.947	0.326	1.57	2.82×10^{-3}
Owner time	-0.194	-0.411	2.38×10^{-2}	8.11×10^{-2}
Recommended by breeder	3.65×10^{-2}	-0.179	0.252	0.740
Owner ability	0.250	-0.133	0.632	0.201
Dog problem	-2.28	-0.341	0.296	0.888
Daily food quantity (g)	5.73×10^{-4}	9.89×10^{-5}	1.05×10^{-3}	1.79×10^{-2}

†Other dog purpose included show, breeding, multi-purpose and all 'other' dogs.

Assistance dogs were excluded because they typically left the project at one year.

4.3.11 Exercise

There were 134,076 lines of exercise information which became 112,060 lines when duplicates were removed. The number dropped further to 97,638 lines when potential owner corrections were taken into account. These 97,638 lines should have been associated with 16,273 data entries or completed questionnaires but a database oddity meant that there were 45 DEIDs that had too much information. These DEIDs were associated with more than one set of questionnaire answers. Unique DEIDs did not apparently correspond to uniquely answered questionnaires. Half of the exercise information associated with 33 of the 45 were deleted because they were duplicated elsewhere. The remaining 12 were retained in complete form but not automatically linked to other information via the DEID. Instead, any such links had to be made by hand.

Each of the 16,000+ different exercise entries, described six different activities at two different time points; weekdays and weekends. As Figure 4.32 demonstrates, there was little difference between weekday and weekend activity levels. If time spent on an activity changed then typically it increased at weekends. Of 97,440 activity reports, 3,963 involved spending less time on that activity at the weekend, 13,771 reports were of more time spent at the weekend and the remainder were unchanged. Only 47.5% (95% CI: 46.7 - 48.3%) of complete exercise regimes did not change at all between the weekday and weekend. An aggregated categorical measure of weekday and weekend activity was created using weighting (as described in Section 4.2.9) and only 4.7% (95% CI: 4.6 - 4.9%) of times spent were different from the weekday report.

The individual points in Figure 4.33 were randomly generated to create continuous data that fit within the given time categories for all entries. The distributions of times spent were strongly right-skewed for all of the exercise types as shown by the higher density of points lower down the plot. The figure also demonstrates the relative popularity, in terms of time spent, of each of the exercise types across the cohort.

The aggregated exercise levels were split according to the age of the dog at the time of reporting and box-plotted in Figure 4.34. Time spent on two categories, 'off lead' and 'fetching, chasing or retrieving', appears to increase as the dogs age but the remaining categories look relatively stable.

The total time spent exercising according to each questionnaire answer was

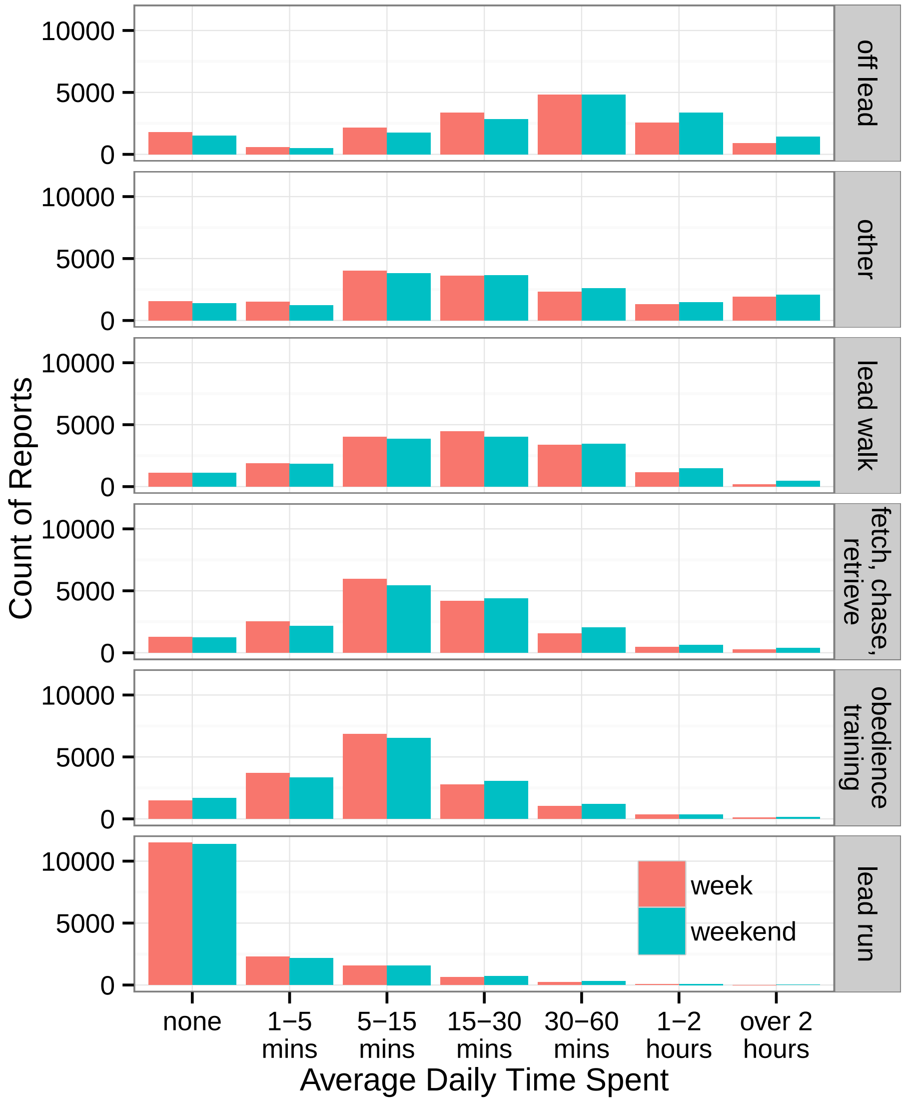


Figure 4.32: Average daily time spent on exercise activities

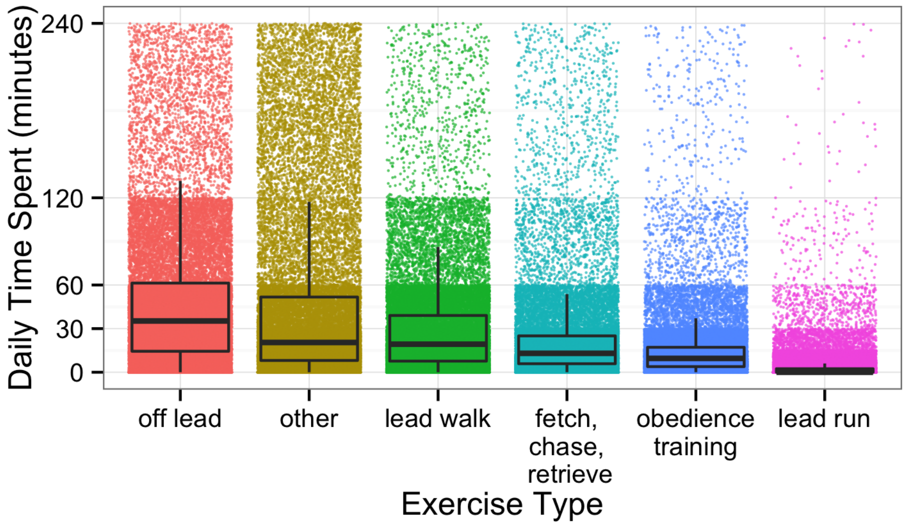


Figure 4.33: Daily time spent on exercise activities

generated using the mid-point of each time category for each exercise type. The distribution of these times was strongly right skewed with mean 157.5 minutes, median 128.7 minutes and range 0 - 18 hours daily (Figure 4.35). The maximum seems rather extreme and was likely either a mistake in data entry or an artefact of assuming that the mid-point of the 'over 2 hours' time category was three hours.

The total times spent exercising were square root transformed to approximately normalise the distribution (Figure 4.36) and correlations were sought with other reported data. Figure 4.37 shows the monthly means of the square root of the time spent exercising each month with 95% confidence bars. The values have been re-squared for ease of interpretation. There appears to be some seasonal variation and if the months are grouped into seasons then it becomes apparent that there is a drop in time spent exercising in autumn (September, October, November), particularly when compared with spring (March, April, May) (Figure 4.38).

Given the seasonality of births (Figure 4.7) and the association between age and total daily time spent exercising (Figure 4.34), it was also important to assess whether the apparent seasonal variation in time spent exercising was due to the changing age profile of the cohort. The plots were repeated using exercise entries made after the dogs became one year of age (Figure 4.39) and the seasonal pattern

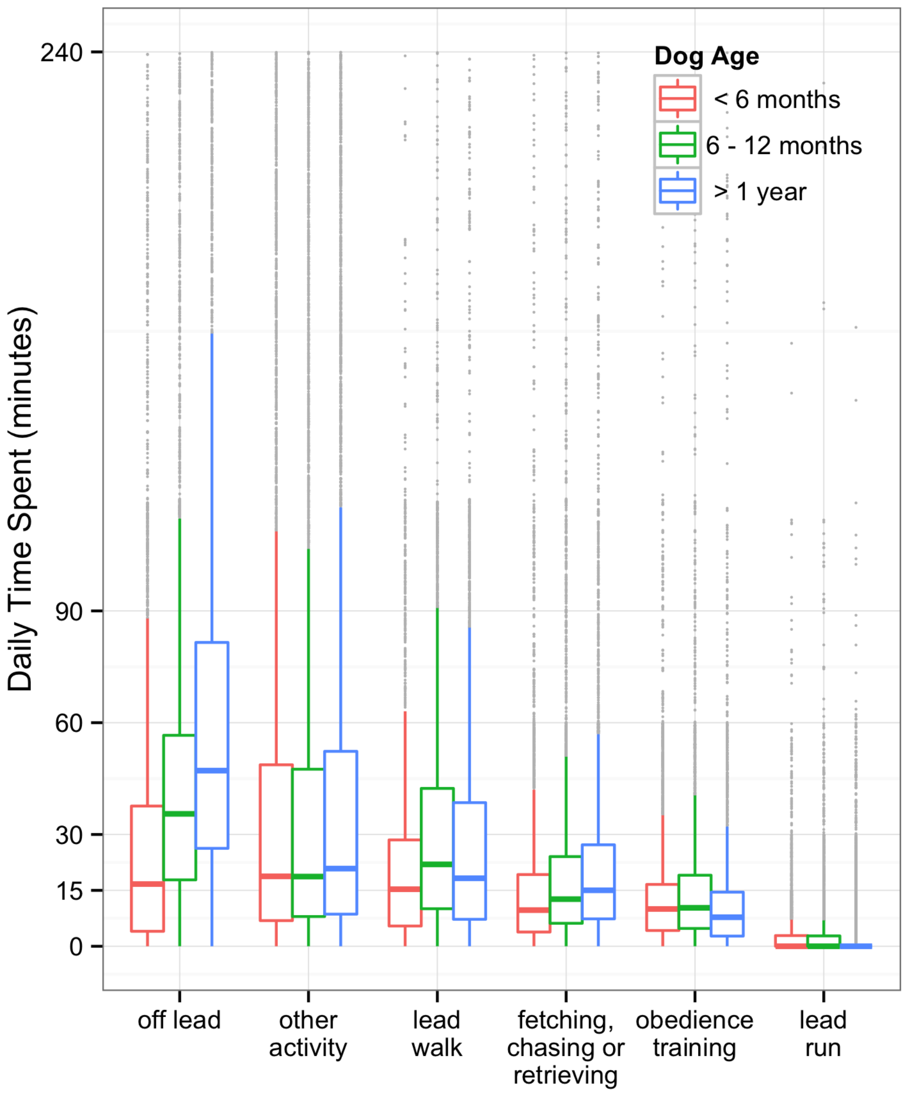


Figure 4.34: Boxplot of daily time spent on exercise activities according to dog age

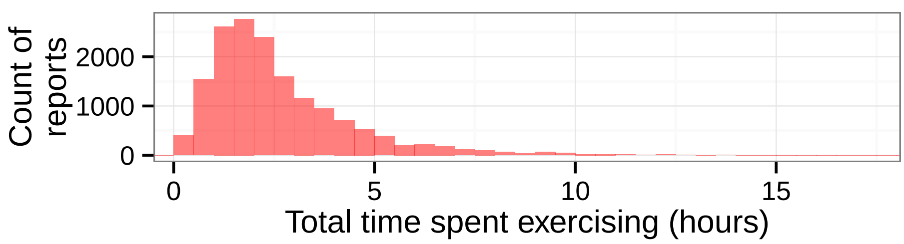


Figure 4.35: Cohort's daily times spent exercising

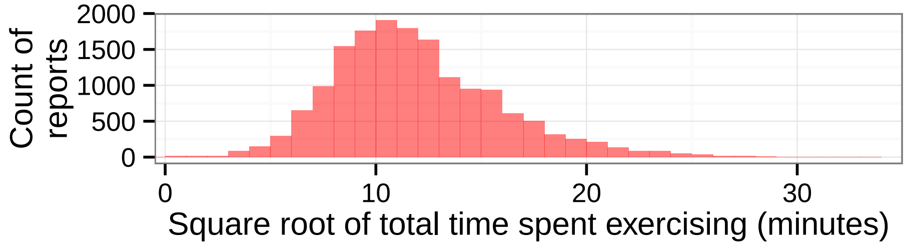


Figure 4.36: Cohort's daily times spent exercising (square root transformed)

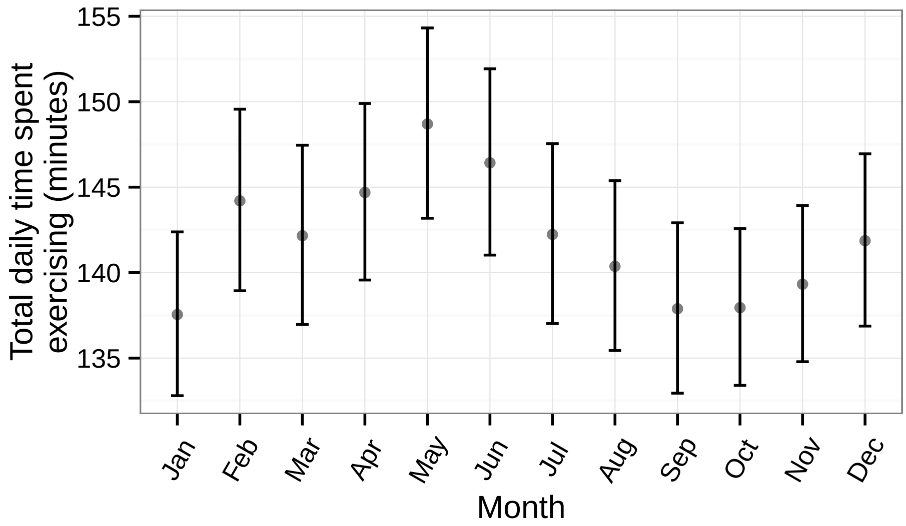


Figure 4.37: Cohort's daily time spent exercising (square root transformed then re-squared) each month. Group mean with 95% confidence bars

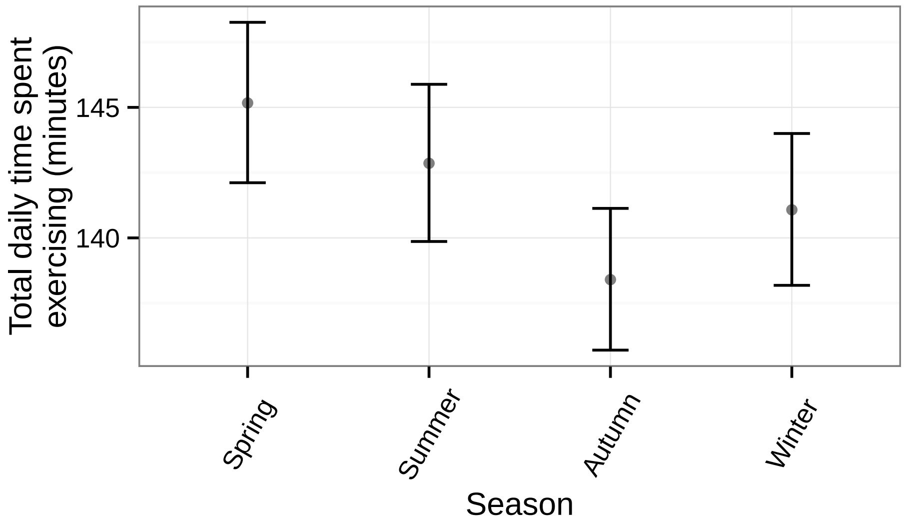


Figure 4.38: Seasonal variation in the cohort's daily time spent exercising (square root transformed then re-squared). Group mean with 95% confidence bars

changed. Winter became the season with the least time spent exercising, though there was considerable overlap between CIs. One might imagine that if variations in exercise amounts were influenced by temperature, or climate more generally, then there would be an interaction between season and location. For example dogs in Scotland, where it is colder and darker than the rest of the UK in the winter months, might be exercised less but there was no apparent interaction between season and country (data not shown).

Independently of season, there were variations in time spent exercising for the different islands and countries where the cohort were located. Figure 4.40 shows averages for each country with 95% confidence bars. The regions that contribute a smaller number of dogs have very large CIs but it is still possible to say that dogs on both Jersey and Guernsey spent less time exercising than the dogs in the other locations. For the larger contributors, Welsh dogs were exercised more than Scottish, and English dogs were exercised less than those in Wales, Scotland or NI.

Figure 4.41 shows the average of the squared root of daily time spent exercising for the different types of dog purposes. Again there is overlap between some 95% CIs but the working dogs spent more time exercising than the pets. This was also

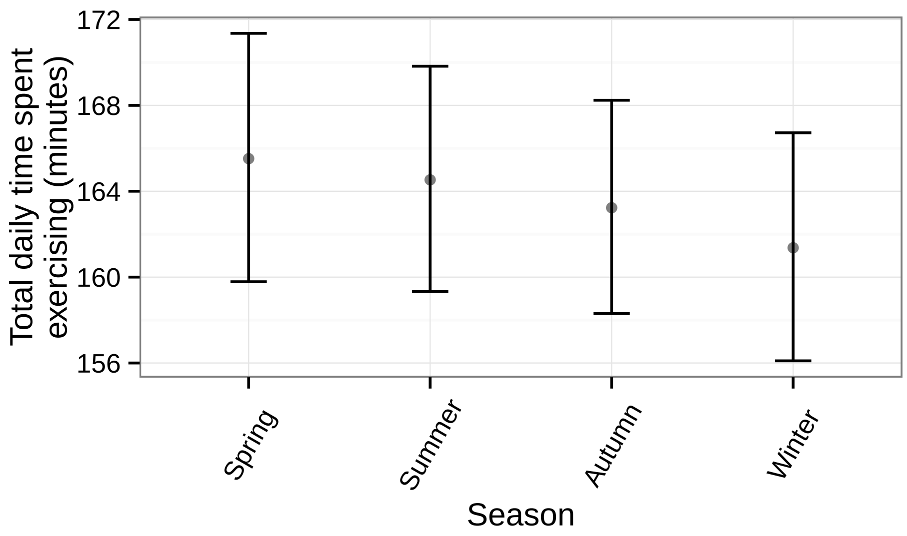


Figure 4.39: Seasonal variation in the cohort's daily time spent exercising (square root transformed then re-squared). Dogs aged one year and over. Group mean with 95% confidence bars

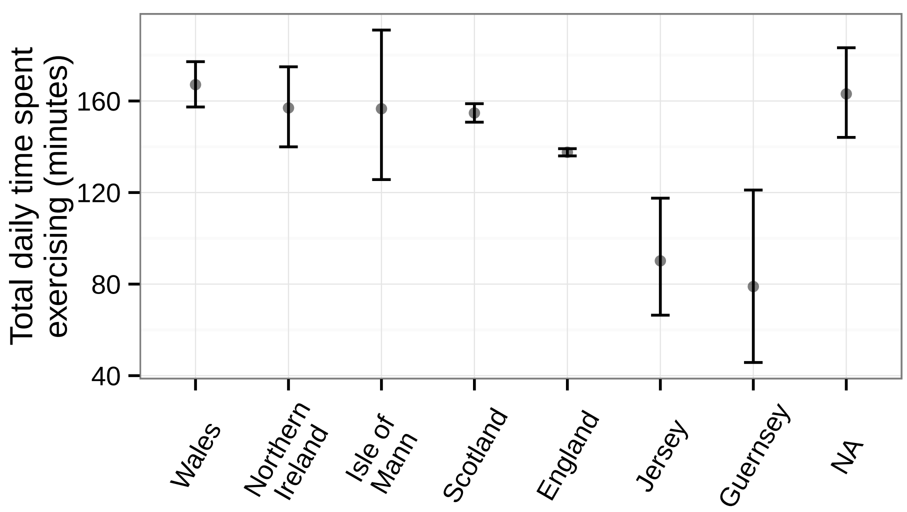


Figure 4.40: Variation in the cohort's daily time spent exercising (square root transformed then re-squared) by location. Group mean with 95% confidence bars

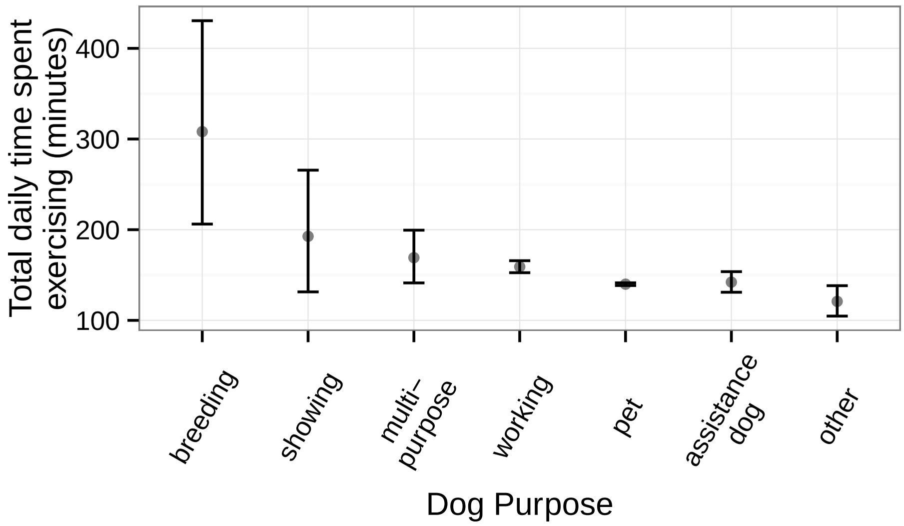


Figure 4.41: Variation in the cohort's daily time spent exercising (square root transformed then re-squared) according to their different purposes. Group mean with 95% confidence bars

reflected in the medians of the original un-transformed times; the median time exercising for working dogs was 138 minutes compared to 128.5 minutes for pets.

Figure 4.42 shows that owner perceptions of restrictions on their dogs' activity levels were reflected in different amounts of time spent exercising. Dogs with unrestricted exercise spent a median of 148 minutes exercising each day which was 14 minutes more than the next group which was restricted due to the location of the household. Dogs with problems that limited their exercise spent just a median of 65 minutes and mean of 89.5 minutes exercising each day. The group of dogs whose owners described their exercise as limited because "recommended by breeder or from their own experience" were also younger on average (mean = 257.0 days, 95% CI: 250.7 - 263.2 days) than the rest of the cohort (mean = 403.7 days, 95% CI: 398.3 - 409.1 days). Figure 4.34 demonstrated a relationship between age and time spent on particular exercise types and it is possible that the relatively low time spent exercising for this group was due to owner restrictions based on dog age.

Figure 4.43 shows that dogs in families were exercised for the lowest mean time each day and dogs in households whose owners did not report a household type were exercised for the longest mean time daily.

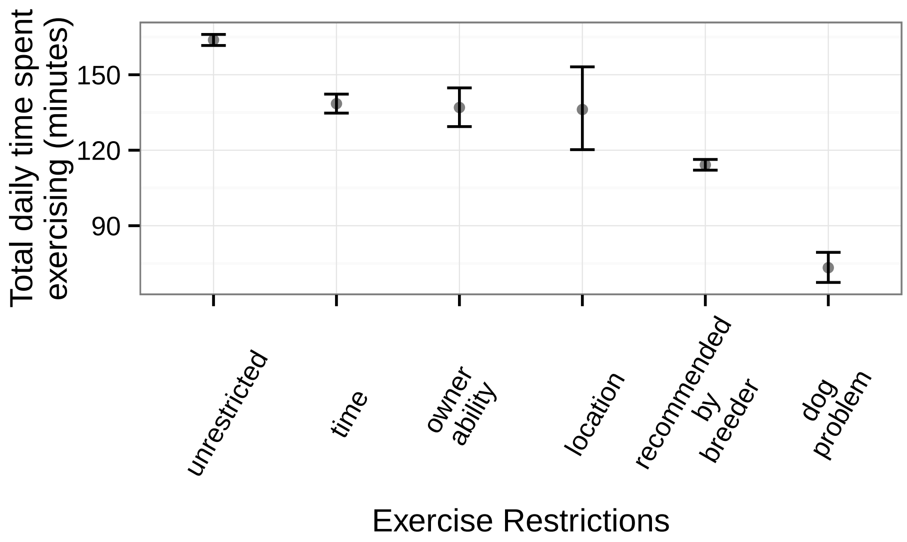


Figure 4.42: Variation in the cohort's daily times spent exercising (square root transformed then re-squared) depending on reasons for exercise restriction. Group mean with 95% confidence bars

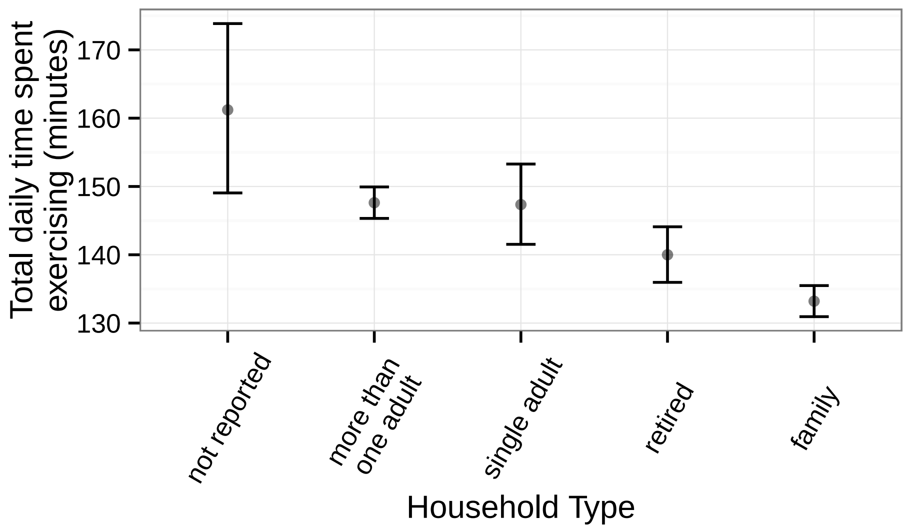


Figure 4.43: Variation in the cohort's daily time spent exercising (square root transformed then re-squared) depending on household type. Group mean with 95% confidence bars

The results of the linear mixed model of all of these factors are shown in Table 4.8. Dogs whose exercise was restricted according to the breeder's recommendations or according to the owner's experience did spend less time exercising than those that were unrestricted, irrespective of the effect of dog age. Dog sex was examined but was there was no obvious different in exercise levels between male and female dogs.

Table 4.8: Linear mixed model of factors associated with total daily time spent exercising

	Coefficient	Std.Error	t value	P value
Intercept	118.23	5.92	19.98	<0.0001
Age (months)	1.42	0.08	18.89	<0.0001
Season				
Spring	0	-	-	-
Summer	-3.49	1.67	-2.09	0.04
Autumn	-5.65	1.64	-3.45	<0.0001
Winter	-4.95	1.66	-2.98	<0.0001
Dog Purpose				
Pet	0	-	-	-
Working dog	15.42	6.04	2.55	0.01
Breed, show or multi-purpose	32.82	14.74	2.23	0.03
Assistance dog	15.87	15.96	0.99	0.32
Other	-26.78	19.58	-1.37	0.17
Location				
England	0	-	-	-
Wales	30.09	8.88	3.39	<0.0001
Scotland	11.42	4.59	2.48	0.01
Northern Ireland	15.83	13.53	1.17	0.24
Isle of Mann	31.92	32.29	0.99	0.32
Jersey	-18.35	54.99	-0.33	0.74
Guernsey	-57.40	93.40	-0.61	0.54
Not given	-0.40	22.26	-0.02	0.99
Household Type				
Family	0	-	-	-
More than one adult	12.71	3.52	3.61	<0.0001
Single adult	20.58	7.49	2.75	0.01
Retired	-4.98	6.33	-0.79	0.43
Not given	32.29	18.48	1.75	0.08
Exercise Restrictions				
Unrestricted	0	-	-	-
Recommended by breeder	-32.14	1.81	-17.75	<0.0001
Time	-12.83	2.57	-4.98	<0.0001
Owner ability	-20.40	4.97	-4.10	<0.0001
Dog problem	-81.91	3.78	-21.69	<0.0001
Location	-18.45	8.50	-2.17	0.03

4.3.12 Sleeping Locations

There were 18,133 lines of data relating to sleeping location. Removal of duplicates and entries that owners themselves amended reduced this number to 16,461. These data comprised 1,063 locations which needed to be grouped. Grouping the answers according to whether they were given before (57.8%) or after (42.2%) the change in questionnaire options allowed for assessment of whether the new options reduced the number of free-text answers. Prior to the change, 18.7% of the answers were given using the free-text option and there were 945 unique answers. After the additional options were offered, just 3.2% of answers were given using the free-text option comprising 155 unique answers. The proportion of free-text answers was significantly reduced ($\chi^2 = 899$, 1 df, $P < 2.2 \times 10^{-16}$). As expected following the validation work (Section 3.3.2.1), the majority of the early free-text answers referred to sleeping with other dogs. Later free-text answers were reduced to those that related to dogs that slept in cages or crates; those that did not sleep in rooms in the house but instead conservatories, hallways, garages, dog rooms etcetera and those whose sleeping routine varied. All answers were fitted into the most appropriate of the revised categories. Those that spent any time sleeping with a person were grouped into the 'with person' category and those that spent any time outside were grouped into the 'outside' category.

The frequencies of reported sleeping locations are shown in Table 4.9. As one might expect from the results of the validation (Section 3.3.2.1), the proportions of reports that described dogs as sleeping alone and with another pet shifted after the question answer options were changed; people no longer had to describe sleeping with another pet in a free-text box.

If the subset of answers given after the change of question is examined, there is a trend that as the dogs aged, the proportion of reports of dogs sleeping alone reduced and in turn, more dogs were reported to sleep with other pets (Figure 4.44). What is not clear from this figure is whether the trend is due to changing owner behaviour as individual dogs age or whether it is due to a different demographic in the owners who stay in the project. As Table 2.5 illustrated, households that included other dogs were *more* likely to be retained in this project which could contribute to the shift in answers.

If only those dogs that contributed more than one data entry on sleeping location are considered, (1,481 dogs after the question change), then 1,129 (76.2%)

Table 4.9: Reported sleeping locations split according to the change in question answer options

Location	Before Question Change (%)	After Question Change (%)	Combined (%)
Alone	5604	3498	
	58.9	*50.4*	*55.3*
With person*	1976	1523	
	20.8	*21.9*	*21.3*
Outside	451	253	
	4.7	*3.6*	*4.3*
With pet	1490	1665	
	15.7	*24.0*	*19.2*

* Includes dogs that sleep with pet *and* person

were not reported to change their sleeping location. The remaining 262 all shifted between categories but not in a fashion that might be expected from Figure 4.44. Indeed more moved from sharing with another pet to being alone (57 dogs) than vice versa (40 dogs) but this is not significant at a 5% level ($\chi^2 = 3$, 1df, $P = 0.08$). The shift in location with age group therefore seems more likely due to bias rather than a true effect.

The total number of dogs reported to sleep outside at least once was 166 and there was some seasonal variation in reporting this location. There were yearly peaks in August in 2011 and 2012, then July in 2013. These dogs were also not split proportionally between the different countries. A relatively high proportion of the dogs that slept outside were from NI, 6.0% (95% CI: 2.9 - 10.8), considering that just 1.5% (95% CI: 1.1 - 1.9%) of the cohort who gave sleeping location data were from NI.

There was a clear association between dog purpose and sleeping location and this was independent of country (country and dog purpose were not correlated). For example, 5.1% (95% CI: 4.4 - 5.9%) of dogs that had sleeping location data were reported to have slept outside at least once but this jumped to 93.2% (95% CI: 88.1 - 96.5%) for working dogs; an odds ratio of 163 (Fisher's test, $P <$ 2.2e-16). No assistance dogs were reported to have slept outside.

Dogs that had been reported to sleep alone at least once were also dispro-

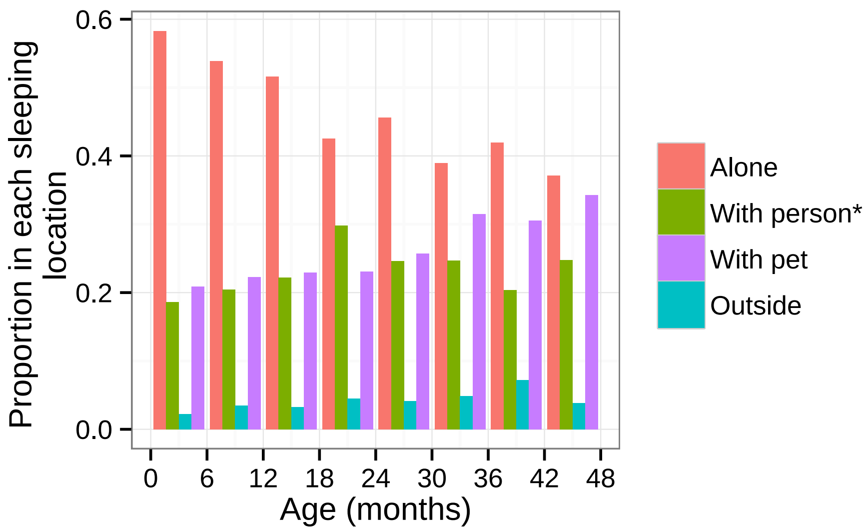


Figure 4.44: Proportion of reports, per age group, for each sleeping location
* With person includes with pet *and* person

portionately split between dogs with different purposes; 36.0% (95% CI: 30.1 - 42.2%) of working dogs were reported to sleep alone at least once compared to 67.0% (95% CI: 65.4 - 68.6%) for the cohort as a whole.

4.4 Discussion

Engaging thousands of dog owners in the Dogslife project has generated a wealth of data that begin to address knowledge gaps regarding UK LR morphology and lifestyle. Selection bias was addressed in Section 2.4 but retention bias remains a potential issue. Table 2.5 indicated that families were more likely to be lost to the project and it is possible that time constraints were a contributing factor because families were also the group who spent the least total time exercising their dogs.

Of the data reported in this chapter, sleeping locations and the proportions neutered were the measures that appeared to be adversely affected by retention bias. For dogs whose owners ever answered the neutering question, just 28.1% of dogs were apparently neutered however this includes many dogs whose owners were effectively lost to the project before their dogs were old enough to be neutered. One would expect the prevalence of neutered dogs in the cohort to increase with age, as shown in Figure 4.16, but the prevalence of neutering in Dogslife dogs over three years of age had reached 0.67 for females and 0.55 for males. These values are considerably higher than 0.41 which was reported in recent work using the veterinary records of 148,741 dogs in the UK (O'Neill et al. [2014b]). This may reflect the differences between Dogslife's population of UKKC registered pedigree dogs and the more mixed group examined by O'Neill et al. but may also indicate that owners who neuter their dogs were more likely to remain in the Dogslife study.

In terms of lifestyle factors, there was considerable homogeneity in the cohort. The majority ate dried food and slept alone. Individual dogs typically did not change diet type but the number of meals per day decreased as the dogs aged. The sleeping location reports highlighted a potential cultural difference between NI and the rest of the UK, with a higher proportion of dogs in NI sleeping outside at least once. NI had a similar mean temperature to both England and Wales in 2013 but had fewer hours of sunshine and more rain (Met Office [2014]) so this was unlikely to be associated with better climatic conditions. The association was found irrespective of dog purpose. From a human perspective, it was interesting that over 20% of reports involved the dog sleeping in the same room as a person. Sensitisation to inhaled dog allergens is one of the major risk factors for asthma Custovic and Simpson [2012] so this may have implications for the health of the owners.

Multiple factors were associated with the total daily time spent exercising. The exercise times of breeding, showing and multi-purpose dogs, and those located in Jersey, Guernsey and the Isle of Mann were based on too few dogs to draw sensible conclusions. Of the four largest regional contributors to the cohort, dogs in England spent less time exercising than dogs in Wales, NI or Scotland. Unsurprisingly, working dogs spent more time exercising than pets, and dogs whose owners reported that their exercise was restricted spent less time exercising than those whose exercise was unrestricted. The clearest difference was for dogs that had a problem, but owners that followed breeder recommendations also spent less time exercising their dogs. This latter type of exercise restriction was associated with younger dogs; younger dogs specifically spent less time 'off lead' and 'fetching, chasing and retrieving'. It could be hypothesised that the dogs were still learning to return to their owners when unrestricted or that breeders advised limiting exercise while the dogs were young because of perceived deleterious effects on musculoskeletal health. Such perceptions can be exemplified by advice from the UKKC (The Kennel Club [2014]).

4.4.1 Dog Weights

Nearly 30 years ago, LR were identified as the most likely breed to be overweight in the UK vet visiting dog population (Edney and Smith [1986]) and it is of concern that the average weight of the cohort increased, linearly, at 0.89 kg per year between one and four years of age. Whilst it is not possible to extrapolate beyond the age range of the data, if this observed increase continues, an expanding proportion of the cohort will become subject to the health consequences of obesity. For example, it has been demonstrated in Elkhounds that there is an association between dogs that were overweight throughout their lives and diabetes mellitus (Wejdmark et al. [2011]) and in LR, an association between higher body weight and increased prevalence and severity of hip dysplasia (Smith et al. [2006]).

The weight model included some surprising results such as chocolate coloured LR being, on average, 1.39 kg heavier than their yellow and black counterparts. Unfortunately there were no yellow dogs homozygous or compound heterozygous for the chocolate alleles at *TYRP1* so it was no possible to ascertain whether this genetic region was contributing to the weight phenotype. Also neutering apparently had no association with weight but a closer look revealed a picture that changed with age. Only after the dogs reached three years of age did the weights of neutered dogs become greater than that of entire dogs and there were not enough dogs of this age to affect the model parameters.

4.4.2 Dog Heights

In 2008, Sutter et al. collected measurements for 1,155 dogs including 14 LR and assessed the percentage of those measured that met the American Kennel Club (AKC) breed standards (American Kennel Club [2014]). It was concluded that the AKC breed standards were a good proxy for height at the shoulder. There is greater allowance for variation in the AKC standard for LR (5.08 cm for each sex in the USA compared to 1 cm for each sex in the UK) but there was also potential for bias in their study. The majority of their sample comprised dogs that had been entered in conformational competitions whereas few of the Dogslife cohort were show dogs. The issue of incorrect measurement or reporting must be considered with all Dogslife data (the height unit error being an obvious example) but visits to a sample of the cohort found no systematic bias to owner height measurements (Section 3.3.4). Therefore whilst individual measurements might be treated with caution, the model parameters should be a good guide to the heights of the population.

Breed standard heights have been used as group phenotypes in studies as proxies for dog size. It is undoubtedly convenient and minimises the time and expense of data collection from individual dogs. However, the Dogslife results suggest two things: firstly that the breed standard is not necessarily the average height for a breed and secondly, that even if it does represent the average, the variability of morphologies might mean that this average poorly reflects many individuals. Under these circumstances, using the breed standard may not be appropriate and might limit the ability of investigators to find true effects. Studies, such as that by Frischknecht et al. [2013], that use individual dog measurements to characterise a phenotype, should have more scope to identify complex patterns. In this instance, it was possible to find potentially causative mutations associated with dwarfism in LR.

4.5 Conclusion

The morphological detail and lifestyle information collected by the Dogslife project offer a unique insight into the lives of pedigree LRs in the UK. These findings set a baseline for further analysis of the relationship between dog morphology, lifestyle and health. It is hoped that Dogslife will contribute to an evidence-based approach to healthy dog ageing.

Chapter 5

Health

5.1 Introduction

Traditionally, canine health data are collected at primary and secondary veterinary facilities. Case notes are kept as part of providing care and it is parsimonious to make use of them for epidemiological studies. However, as discussed in Section 1.1, data collected at both primary and secondary veterinary centres are subject to known biases and, by definition, exclude signs of potential illness that do not precipitate veterinary visits. Unusually, Dogslife collect data directly from owners and this facilitates the collection of details of problems unreported elsewhere. The illness-related section of the questionnaire (Appendix 2, Section A2.7) starts by asking the owner whether their dog has had any of a list of problems and, only if they answer 'Yes', does the questionnaire go on to ask whether they visited the vet. This distinctive approach offers a greater depth of health information that may be used to investigate disease aetiology. There is the potential that signs which do not precipitate veterinary visits may nevertheless be identified as risk factors for subsequent poor health.

This chapter details the Dogslife illness data collected between 1st July 2010 and 31st December 2013. It enumerates all illness reports and provides an owner's perspective of the disease burden suffered by young, pedigree LRs. The twenty most frequently reported signs are examined in terms of illness rates at different ages, demonstrating how the disease burden changes as the dogs age.

Dogslife have a wealth of data regarding the health of LRs and the database is a unique resource regarding illnesses that would not be recorded in traditional epidemiological studies. These reports were dominated by high numbers of gastrointestinal signs. The potential for in-depth analysis of signs which typically were not associated with veterinary presentation is demonstrated by a detailed analysis of the risk factors associated with vomiting and diarrhoea.

5.2 Methods

The data were examined, cleaned, tabulated and graphed using R (R Core Team [2013]). Specific packages used are given in *italics*. Data cleaning procedures were described in Section 2.3.1.

5.2.1 Illness Coding

Illness reports had different levels of detail. Owners were asked different questions according to whether they visited the vet and whether they took a Dogslife Health Report (DHR) for their vet to complete. Due to changes in the questionnaire following validation, the timing of the report also affected the extent of information collected. The potential information available may be summarised as follows:

- No veterinary visit pre February 2013: Type of problem only.
- No veterinary visit post February 2013: Type of problem and owner diagnosis.
- Veterinary visit but no DHR: Type of problem and veterinary diagnosis (owner recollection).
- Veterinary visit and DHR: Presenting sign(s) and diagnosis (completed by veterinarian, copied into Dogslife by owner).

A veterinarian member of the Dogslife team (DNC) reviewed each potential illness and coded them with presenting sign(s) and diagnoses using the VeNom coding system (*http://www.venomcoding.org*). Diagnosis codes included 'Diagnosis not made' to account for the many instances when information was unavailable or insufficient. The coding was undertaken every 2-4 weeks and took 2-3 hours with approximately 50-60 illnesses coded per hour. When necessary, and where contact was possible, ambiguities were resolved by asking the owner for clarification. By coding relatively contemporaneously, if clarifications were required, the owner could be contacted relatively quickly following an ambiguous questionnaire answer. The illnesses in this chapter are described based on presenting sign codes rather than diagnoses because they are the most comprehensive summary measure of owners' reports.

5.2.2 Missing Illness Dates

In addition to basic information about the type of problem the dog was suffering from, the online questionnaire facilitated the collection of more data about when the problem occurred and how long it lasted. However, all questions presented to the owner after they ticked 'Yes' with regard to a specific sign of potential illness were optional and many entries had missing data.

In the case of missing illness dates, entries were individually viewed to determine whether free-text inputted by the owner might include a mention of dates. If omissions remained with regard to when the sign started, and if the owner had entered an end date for the sign, this was used as a start date. If the entry had neither a start nor an end date but had a veterinary visit date, this was used as a start date. The remaining entries that had no start date were flagged and given a start date of the date of reporting.

5.2.3 Date Cleaning

There was no automatic logic check regarding dates that owners might enter. For example, the website did not flag an error for the owner if they entered a start date for an illness if the date was before the dog was born. Such checks therefore had to be undertaken prior to analysis. The checking process was based on the premise that dates would be changed if there was a single, obvious alternative to an inconsistent date. Guiding this process for illness entries was a dog's date of birth, the entered start, end and veterinary visit dates for a presenting sign, and the date the owner reported the information. There was an element of subjectivity in the process but all amendments were coded and thus available for future review and audit.

5.2.4 Syndrome Coding

One unfortunate artefact of the illness data collection process was the generation of duplicates. An owner could describe one vomiting instance and it might be found in the Dogslife database multiple times, each with a unique (Illness identifier (IID)). Accounting for or removing these duplicates was complicated by the fact that one (or more) of the different IID might be associated with more detailed information regarding veterinary treatment. In raw data terms, there were 7,096 unique IIDs in the Dogslife database on 1st January 2014 but this equated to fewer reported signs because a number of those IIDs were artificially created duplicates. The 7,096 entries comprised 7,359 rows of data because, for example, a free text entry of "gummy ears and eyes" would be coded with two separate presenting signs for the eye and ear signs. Once those coded as "Not presented for a complaint" (n = 36) were disregarded, 7,323 rows of illness data remained for cleaning and analysis (including duplicates).

An aim was to try to distinguish between presenting signs that were reported on their own and those that were reported as co-occurring with other signs (such as the eye and ear problem above). Ideally this would have been done using reported start and end dates but 28.6% (95% CI: 27.6 - 29.7%) of reported signs had no end date. Further, an ongoing limping or lameness episode may have lasted for many months and not be related to an episode of vomiting that occurred during that period. Instead signs were provisionally grouped according to whether they started within three days of each other. This process had the advantage of effectively grouping the artificial duplicates mentioned above into one 'syndrome'. Henceforth, the word syndrome will be used to mean a reported event comprising one or more signs of potential illness. Duplicated signs could then be ignored without removing any associated information such as veterinary diagnoses. The only exceptions to this process were the entries that were flagged as having been given the reporting date as a start date (as described in Section 5.2.2). It was recognised that these start dates were artificial and might falsely group together signs that had been reported at the same time but had not necessarily co-occurred. Instead these entries were only grouped into syndromes if they were coded with multiple presenting signs for one entry (such as the eye and ear problem previously mentioned).

5.2.5 Descriptive Analysis

The number of times presenting signs were reported were described with reference to the frequencies of reporting for individual dogs. Rates of events were also calculated using Therneau's *survival* package (Therneau [2014]) with poisson confidence intervals. Time at risk was divided such that an interval of 3-6 months would include 3.0 months but exclude 6.0 months. Age in months was taken from the age in days multiplied by 12/365.25. Time at risk during individual months was calculated from the length of the relevant month, for example, February in 2012 comprised 29 days.

5.2.6 Modelling

Where time to event data were considered, Cox proportional hazards models (Equation 2.1) (Cox [1972]) were applied using the *survival* package (Therneau [2014]). Models were checked by plotting time-varying estimates of the log of the hazard ratio for each parameter. Where the parameters appeared to change with time, implying non-proportionality, two approaches were tried. Firstly an interaction term was included between the varying parameter and the dog age and secondly the whole model was stratified by dog age. Proportionality was re-checked and if both approaches resolved the issue, the method that maximised the model R^2 value was chosen.

5.2.6.1 Diarrhoea and Vomiting

The number of times diarrhoea and vomiting were reported as presenting signs was described, both on their own and when in combination with other signs. Proportions were presented with binomial confidence intervals. Incident frequency in individual dogs was put in the context of their time at risk. Potential associations were investigated between duration and severity of signs and whether an owner took their dog to the vet. Risk ratios were determined using 2x2 tables of diarrhoea and vomiting events and concurrent exposure to raw diets. Associations were sought between changes in broad diet types and reports of diarrhoea and vomiting. Cohort-wide cumulative incidence was determined and incidence rates at different ages were presented graphically. Due to the possible impact of recall decay on illness reporting (Sections 3.6.4 and 4.3.2), incidence rates were assessed using the whole cohort time at risk and the more restricted time at risk generated by only considering a window of 40 days prior to each data entry. This process was described in Section 4.2.2 with the aid of Figure 4.2.

Incidence rates for diarrhoea and vomiting were mapped across UK postcode areas and modelled across postcode districts using the *maptools* package in *R* (Bivand and Lewin-Koh [2015]). For example, the Roslin Institute postcode is EH25 9RG and it would fit in postcode area EH and district EH25. Latitude and longitude were available for all dogs that had associated postcodes and associations between illnesses and latitude and longitude were assessed in univariable and multivariable models. Human population density was available for postcode districts in Scotland, England and Wales but not NI or many islands. Human population density was calculated as the number of people in the 2011 census (Office for Naitonal Statistics [2011] for England and Wales and National Records of Scotland [2011] for Scotland) divided by the area of the district in hectares and reported in 100s of people per km^2. Where population density was significant in multivariable models, only Great Britain (comprising England, Wales and Scotland) could be considered.

5.3 Results

5.3.1 Missing dates and date cleaning

Optional date fields relating to illnesses included start, end and veterinary visit dates. Veterinary visit dates were the most complete. Owners reported that they took their dog to the vet in 3,487 entries and of these, only 117 had no associated veterinary visit date. At the time of data entry, owners reported that the sign of illness had ended in 5,444 cases and of these, there were 273 without end dates. If the sign was ongoing at the time of reporting, it was not possible to enter an end date. There were 238 entries with no start date.

The entries with no start date were individually examined. Fifteen entries could be given either a start date or a veterinary visit date from other available data and a further eight could be deleted as the same sign was reported at the same time but with start, end or veterinary visit date included. Three signs were deleted because the owner had given no description or dates. As described in the methods, Section 5.2.2, the end date was used as the start date in 17 instances

Table 5.1: Errors identified in the illness dates

Error Type	Reason for Error	Action
Start, end or	Out-by-year	Amended 12
veterinary visit	Out-by-month	Amended 3
dates before DOB	Out-by-year and month	Amended 7
End or veterinary	Out-by-year	Amended 6
visit dates	Out-by-month	Amended 12
before start date	Typo	Amended 9
	Unclear	Unchanged 56
Start or end	Out-by-year	Amended 7
dates after date	Out-by-month	Amended 170
of owner report	Out-by-year and month	Amended 17
	Unclear	Unchanged 62

and the veterinary visit date was used in 32 instances. The remaining 167 entries were flagged and given a start date of the date of reporting.

Checks regarding reported illness dates identified 391 that were illogical. The errors identified and actions taken to resolve them are shown in Table 5.1.

5.3.2 Syndromes

There were 305 unique syndromes created by grouping signs according to their start dates. The syndrome list was examined by a veterinarian (DNC) for biological plausibility and 75 were identified as unlikely to be co-occurring due to the same process. For example, Faecal appearance abnormal - *diarrhoea* and Pruritus were reported together 111 times but it was not thought that they could be linked biologically. Each instance of these 'non-syndromes' was scrutinised and 13 of 177 entries were determined to be plausible syndromes. The remaining 164 entries were split back up into individual signs or groups of signs that could be plausibly linked biologically. Once unlikely groupings were removed, 255 unique syndromes remained comprised of 104 different VeNom coded presenting signs. The maximum number of different signs in a syndrome was five. The syndromes varied in frequency with a total of 6,115 syndromes associated with 2,099 dogs. Of the 6,115 syndromes, 47.4% were associated with a veterinary visit. The median number of syndromes per dog was two with a maximum of 17 (Figure 5.1).

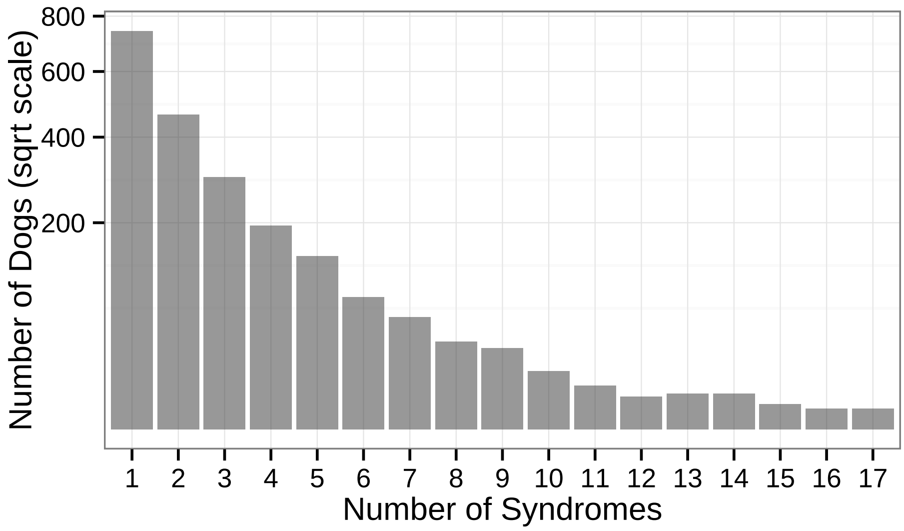


Figure 5.1: Number of Syndromes per Dog

The twenty most frequently reported syndromes, comprising 90% of reports, are shown in Table 5.2. These syndromes have been grouped according to system or event type and when their incidence rates are compared, age-related patterns of reporting become apparent (Figure 5.2). Illnesses were dominated by gastrointestinal signs which peaked between 3-6 months. Musculoskeletal signs peaked later between 6-9 months and wound/trauma peaked later still at 9 months - 1 year. These findings were reported at the 2015 meeting of the Society for Veterinary Epidemiology and Preventative Medicine and published as part of a paper in the conference proceedings (Pugh et al. [2015a]). All syndromes reported to Dogslife are listed in Appendix 6.

5.3.3 Diarrhoea

Diarrhoea was reported 1,622 times; 26.5% (95% CI: 25.4 - 27.7%) of illness reports included diarrhoea. It was associated with 1,111 dogs which is 34.4% (95% CI: 32.8 - 36.1%) of all dogs that had reported data and 52.9% (95% CI: 50.8 - 55.1%) of the dogs that had an illness report. The number of times each dog was reported to have had diarrhoea can be seen in Table 5.3. The vast majority of these reports involved diarrhoea on its own but it was also reported as part of 19 other syndromes (Table 5.4) including 363 reports of diarrhoea co-occurring with vomiting.

Table 5.2: Twenty most frequently reported syndromes

Sign	Sign	Frequency	Proportion involving vet visit	95% CI lower	95% CI upper
[G]Faecal appearance abnormal - *diarrhoea*	none	1215	0.31	0.28	0.34
[G]Vomiting - *other*	none	1094	0.16	0.14	0.19
[S]Pruritus	none	931	0.34	0.31	0.37
[M]Gait abnormality - *lameness*	none	792	0.66	0.63	0.70
[G]Faecal appearance abnormal - *diarrhoea*	Vomiting - *other*	363	0.47	0.42	0.53
[R]Coughing	none	282	0.70	0.65	0.75
[Ey]Ophthalmic (eye) abnormality	none	160	0.86	0.79	0.91
[Ea]Ear (aural) abnormality	none	156	0.81	0.74	0.87
[O]Presenting complaint not listed	none	77	0.56	0.44	0.67
[WT]Wound	none	71	0.77	0.66	0.87
[S]Skin (cutaneous) abnormality - *other*	none	57	0.74	0.60	0.84
[S]Mass/swelling - *skin (cutaneous)*	none	56	0.88	0.76	0.95
[Ey]Discharge - *ocular (eye)*	none	41	0.73	0.57	0.86
[S]Skin (cutaneous) abnormality - *eruptions/hives/rash*	none	35	0.86	0.70	0.95
[O]Anal irritation	none	34	0.88	0.73	0.97
[O]Urination abnormal - *other*	none	33	0.76	0.58	0.89
[WT]Traumatic episode	none	31	0.74	0.55	0.88
[O]Mass/swelling - *other*	none	28	0.82	0.63	0.94
[R]Coughing	Vomiting - *other*	25	0.84	0.64	0.95
[M]Musculoskeletal injury	none	21	0.52	0.30	0.74

[G] Gastrointestinal; [S] Skin-related; [R] Respiratory; [M] Musculoskeletal; [Ey] Eye-related; [Ea] Ear-related; [WT] Wound/Trauma; [O] Other

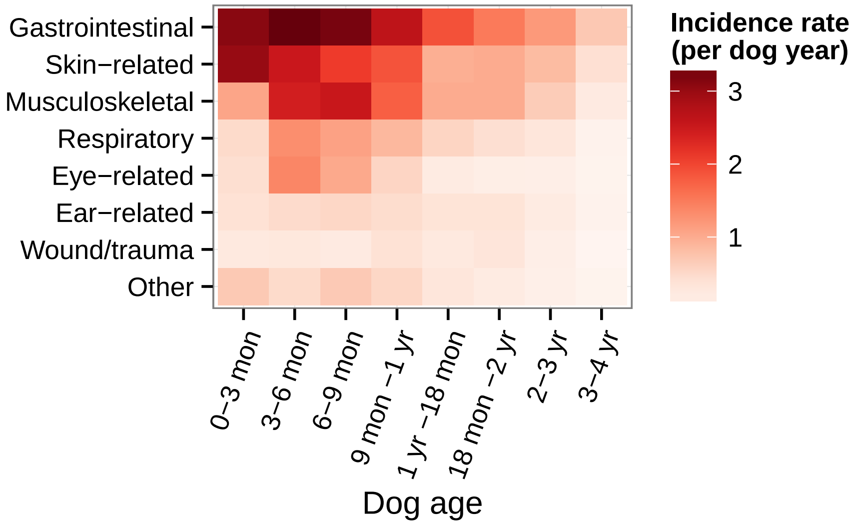


Figure 5.2: Illness rates for twenty most frequently reported syndromes, split according to dog age. The syndromes have been grouped according to system or event type as described in Table 5.2

Table 5.3: Number of dogs that have different reporting frequencies for diarrhoea and vomiting

		Number of reports								
		1	2	3	4	5	6	7	8	10
Diarrhoea	(1,111 dogs)	774	227	75	18	12	6	1	3	0
Vomiting	(997 dogs)	637	183	98	34	14	7	0	3	1

Table 5.4: Other signs reported with diarrhoea and the syndrome frequency

Other sign	Other sign	Freq	Proportion involving vet visit *95% CI*
none	none	1215	0.31 *(0.28 - 0.34)*
Vomiting - *other*	none	363	0.47 *(0.42 - 0.53)*
Dietary indiscretion - *other*	none	13	1.00 *(0.75 - 1.00)*
Faecal appearance abnormal - *haematochezia*	none	5	1.00
Vomiting - *other*	Vomiting - *haematemesis*	5	0.40
Vomiting - *other*	Dietary indiscretion - *other*	3	1.00
Vomiting - *haematemesis*	none	3	0.67
Vomiting - *other*	Dietary indiscretion - *foreign body ingestion*	2	1.00
Presenting complaint not listed	none	2	1.00
Anal irritation	none	1	1.00
Dietary indiscretion - *foreign body ingestion*	none	1	1.00
Faecal appearance abnormal	Vomiting	1	1.00
- *haematochezia*	- *other*	1	1.00
Faecal appearance abnormal	Vomiting	1	1.00
- *haematochezia*	- *haematemesis*	1	1.00
Faecal appearance abnormal - *other*	none	1	1.00
Gait abnormality - lameness	none	1	1.00
Mass/swelling - other	none	1	1.00
Salivation - *increased/drooling*	none	1	1.00
Vomiting - *other*	Lymphadenomegaly	1	1.00
Vomiting - *other*	Pyrexia/hyperthermia	1	1.00
Weight loss	none	1	1.00

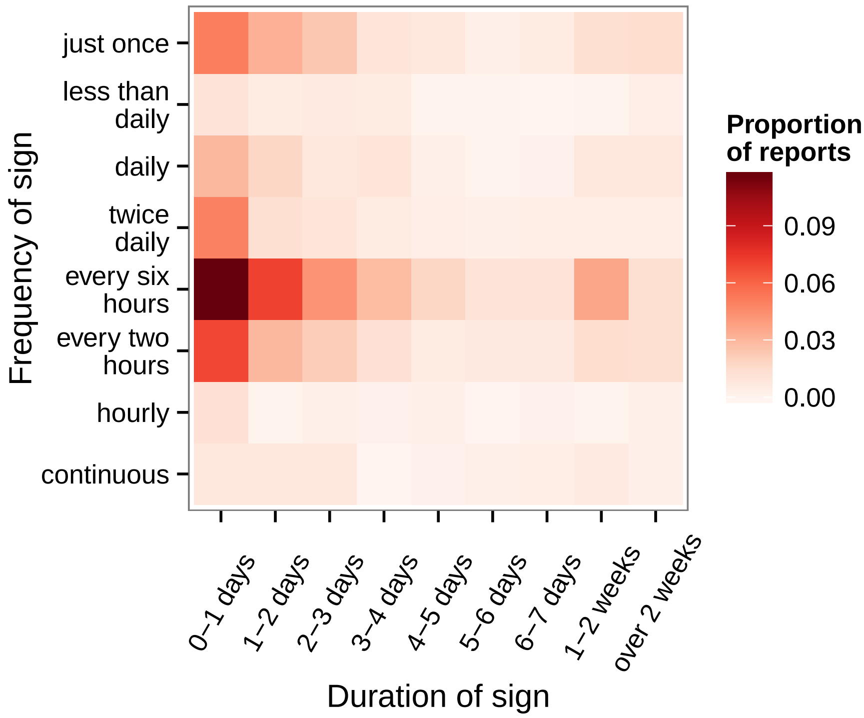


Figure 5.3: Duration and frequency of reports of diarrhoea

No association was found between dogs that were fed raw diets and diarrhoea (risk ratio 0.74, 95% CI: 0.46 - 1.2). In terms of broad diet type, for example 'dried' or 'home prepared', there was no association between change in diet type and diarrhoea (risk ratio 0.87, 95% CI: 0.67 - 1.15).

Of the 1,215 reports of diarrhoea only, 1,068 had both start and end dates. They are shown, broken down according to the frequency and duration of the sign in Figure 5.3. Owners were most likely to report that the diarrhoea occurred every six hours and the majority of incidences lasted less than four days.

The proportions of each of these groups of signs that involved a veterinary visit are shown in Figure 5.4. As might be expected, the likelihood of a veterinary visit increased as the duration increased. Duration appeared to be more important than frequency of sign in terms of the owner decision to take their dog to the veterinarian.

The reports of illness events which included diarrhoea were filtered according to whether they occured within the total cohort time at risk. This reduced the number of reports from 1,622 to 1,560 because 62 incidences took place more than

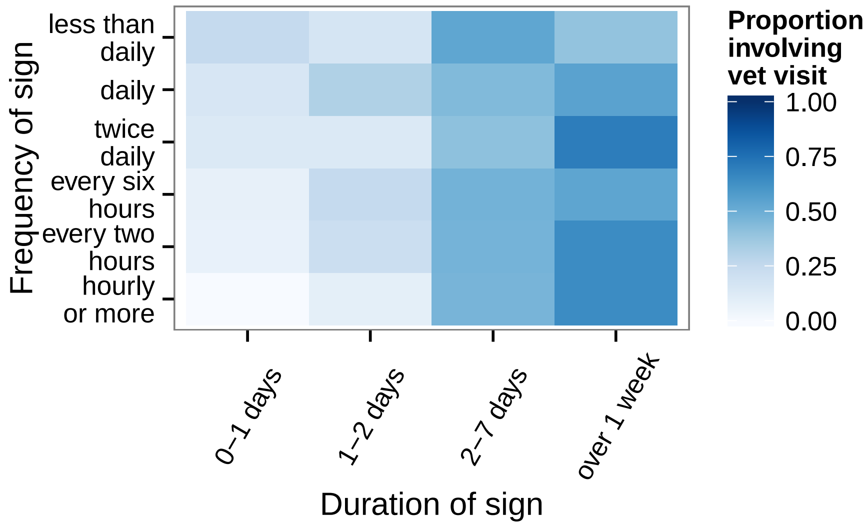


Figure 5.4: Proportion of diarrhoea reports involving a veterinary visit

28 days before the first completed questionnaire. The overall incidence rate for the cohort was 0.50 (95% CI: 0.48 - 0.53) incidents per dog year but the incidence rates decreased with dog age and this is illustrated in Figure 5.5.

Reducing the overall time at risk under consideration to minimise recall decay excluded a further 123 reports, reducing the numbers from 1,560 to 1,437 and the time at risk to 1,733 dog years. The overall incidence rate increased to 0.83 (95% CI: 0.79 - 0.87) incidents per dog year but again, the incidence rate decreased as the dogs got older (Figure 5.5).

The maximum number of reports of diarrhoea for one dog was eight times which was reported for three separate dogs. The remaining frequencies of diarrhoea reports are included in Table 5.3. The majority of dogs that had diarrhoea reports had just one report (774 of 1,111 dogs). As suggested by the findings in Table 3.15, illnesses were under-reported to Dogslife. A UK-based study found the prevalence of diarrhoea in dogs in the preceeding two weeks to be 19% (Hubbard et al. [2007]) indicating that diarrhoea is ubiquitous. The Dogslife records of dogs that had no diarrhoea report were examined to determine whether they simply contributed less time to the project, i.e. they had not had diarrhoea *yet*. There was a difference between the groups in terms of the distribution of time at risk which is shown in Figure 5.6. The bulk of dogs that did not have reports of diarrhoea contributed less time to the project.

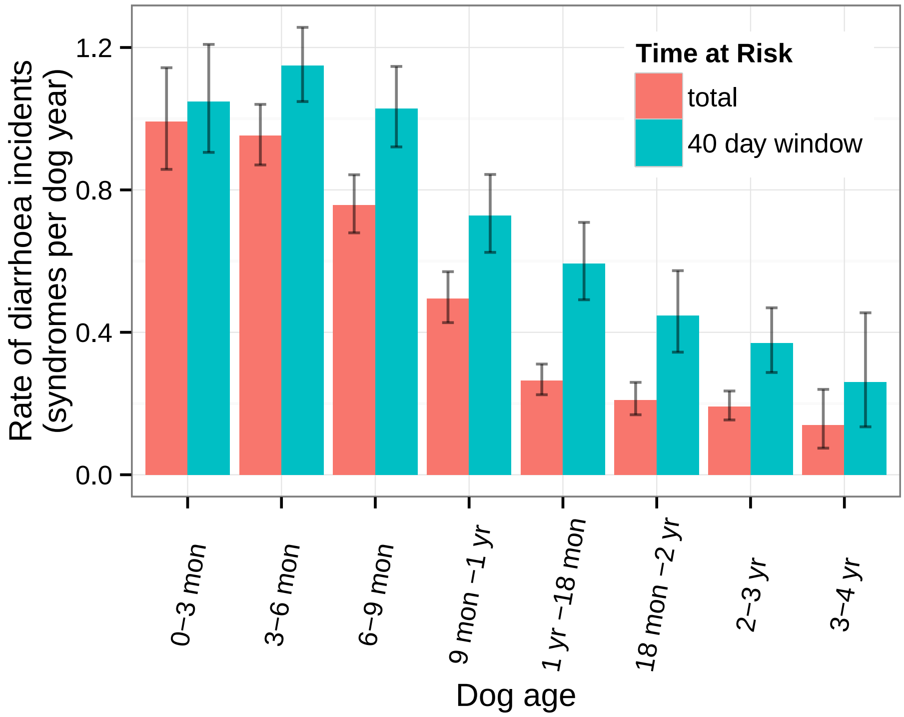


Figure 5.5: Incidence rates of diarrhoea at different dog ages. Total time at risk includes 3,098 dog years and 1,560 reports of diarrhoeal illness dropping to 1,437 reports and 1,733 dog years with windows of 40 days applied. Bars represent 95% poisson CI.

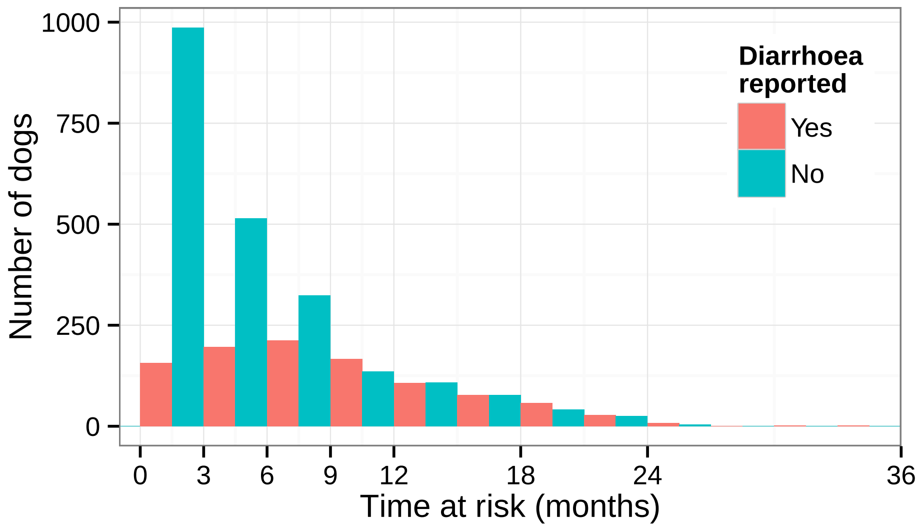


Figure 5.6: Time at risk for dogs that do and do not have reports of diarrhoea. Windows of 40 days applied. Dogs with zero time at risk excluded.

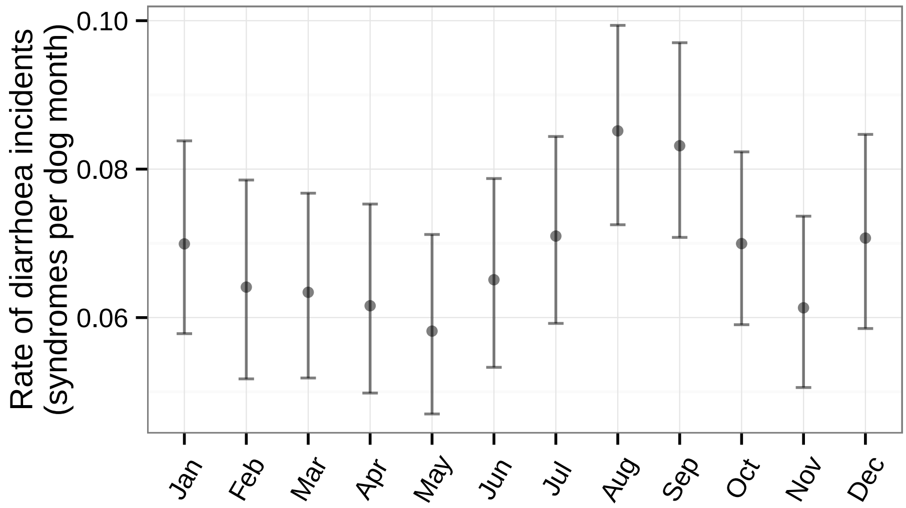


Figure 5.7: Month-specific incidence rates for diarrhoea with 95% poisson CI. Windows of 40 days applied.

Considering just the events and time that were included once the 40 day windows were applied, the monthly incidence rates for diarrhoea were estimated (Figure 5.7). There appeared to be a seasonal pattern but, given the seasonality of births (previously shown in Figure 4.7) and the variation in rates at different ages, it was not clear whether this pattern was due to the high proportion of relatively young dogs contributing to events in August and September each year. Stratifying the time at risk according to the age of the dog when the syndrome first started (Figure 5.8) made it clear that it was not just a reflection of the seasonal birthing pattern. There was still a peak in August and September for dogs under the age of one year, particularly for dogs between three and nine months. This difference was confirmed using a univariable Cox proportional hazards model, assessing the first incident of diarrhoea in dogs under one year of age (Table 5.5).

Postcode area rates of first diarrhoea incidence for each dog were mapped (Figures 5.9 and 5.10) using raw latitude and longitude rather than a specialised projection. Just the first event was considered to concentrate on the between dogs effect. There was a trend towards higher rates in the south and east of the country. The times to these first incidences of reported diarrhoea were explored and the results of univariable Cox proportional hazards models are shown in Table 5.6. Both the total time at risk and the time adjusted to the 40 days prior

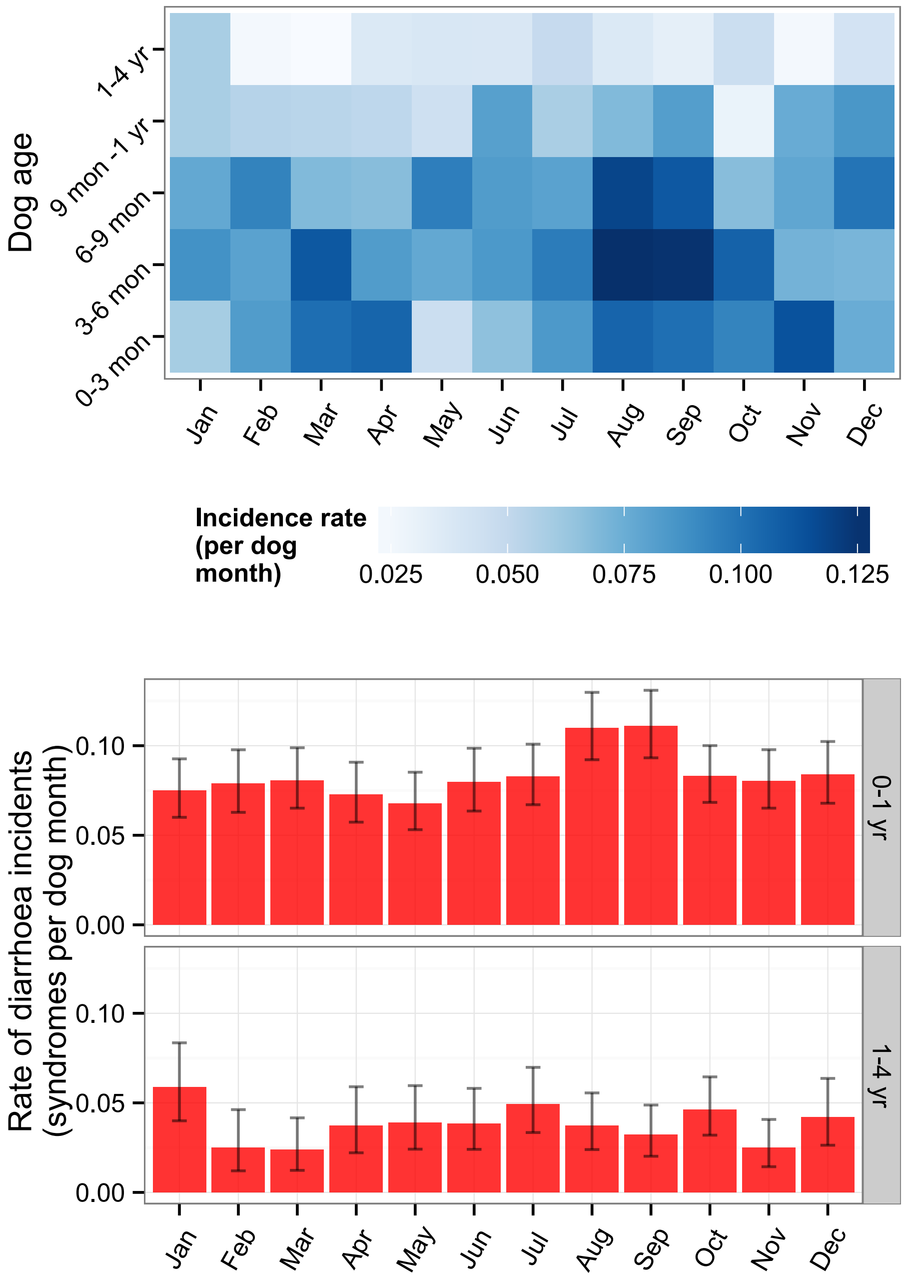


Figure 5.8: Age and month-specific incidence rates for diarrhoea. Windows of 40 days applied.

Table 5.5: Univariable Cox proportional hazards model of association between month of the year and time to first diarrhoea event (dogs under one year only)

Month	Hazard Ratio e^{β}	95% CI: lower	upper	*P*-value
January	1.15	0.791	1.67	0.47
February	1.24	0.854	1.81	0.26
March	1.46	1.02	2.08	**0.038**
April	1.15	0.783	1.69	0.48
May	1	-	-	-
June	1.11	0.757	1.62	0.60
July	1.24	0.866	1.77	0.24
August	1.81	1.30	2.52	<**0.001**
September	1.63	1.16	2.28	**0.004**
October	1.20	0.848	1.71	0.30
November	1.20	0.841	1.72	0.31
December	1.25	0.872	1.80	0.22

to each data entry were considered. For the total time at risk, the median time to the first incident for each dog was 722 days (just under two years) (95% CI: 556 - 984 days) but this halved to just 364 days (95% CI: 327 - 431 days) when the windows of 40 days were applied. The drop was unsurprising because only 58 of 1,071 reports lay outside the windows but applying the windows dropped the time at risk from 2,639 to 1,511 years. The mean times to the first incidents were 771 days (95% CI: 745 - 798) for the total time at risk but just 617 days (95% CI: 582 - 652 days) when the time at risk was adjusted.

The univariable Cox proportional hazards analyses (Table 5.6) confirmed the impression given by Figures 5.9 and 5.10; the rate of diarrhoea events was lower in Scotland than England and dogs in Scotland suffered their first diarrhoea incident later. Rather than using broad country boundaries, it was possible to use latitude, longitude and human population density (Section 5.2.6.1). Decreasing longitude, increasing latitude and lower human population density were all associated with increased time to first report of diarrhoea. Having another dog in the household was also associated with a lower hazard ratio.

Interestingly, associations between household type and time to diarrhoea event differed according to which at-risk period was applied, indicating a potentially differential impact of recall decay on different households. In both cases, the

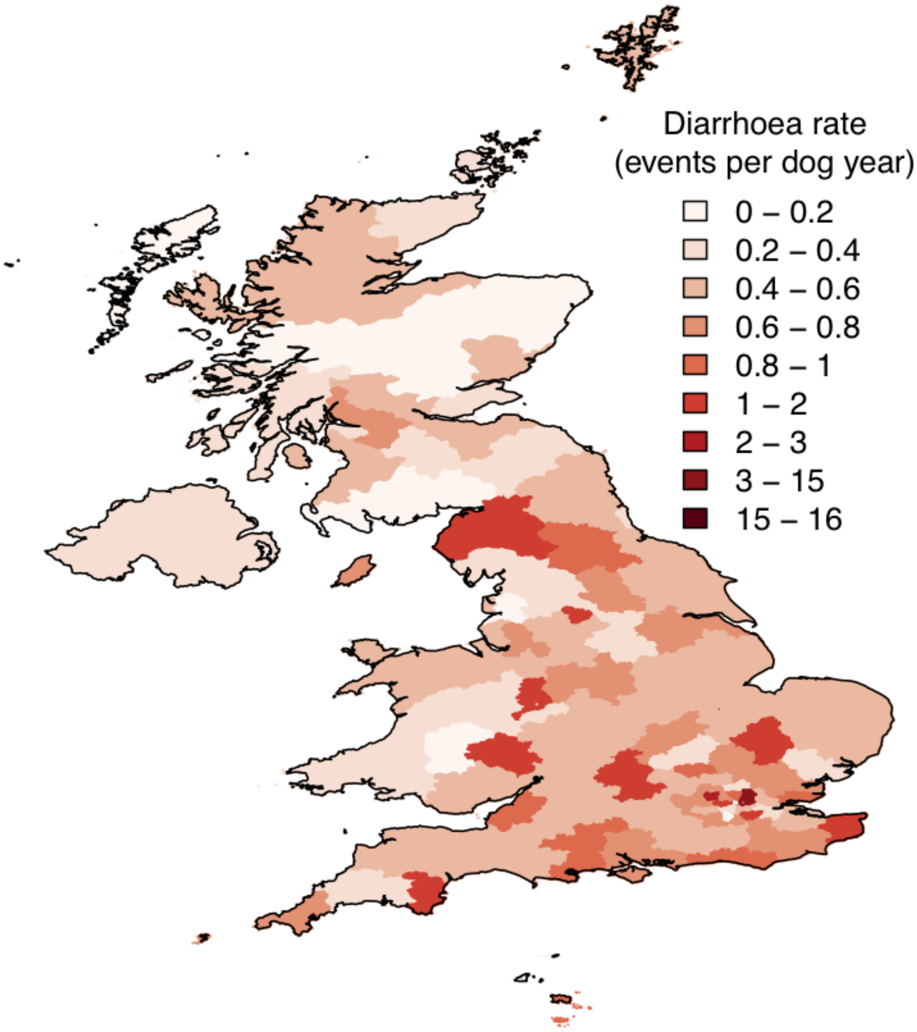


Figure 5.9: Postcode area diarrhoea incidence rates.
Total time at risk.

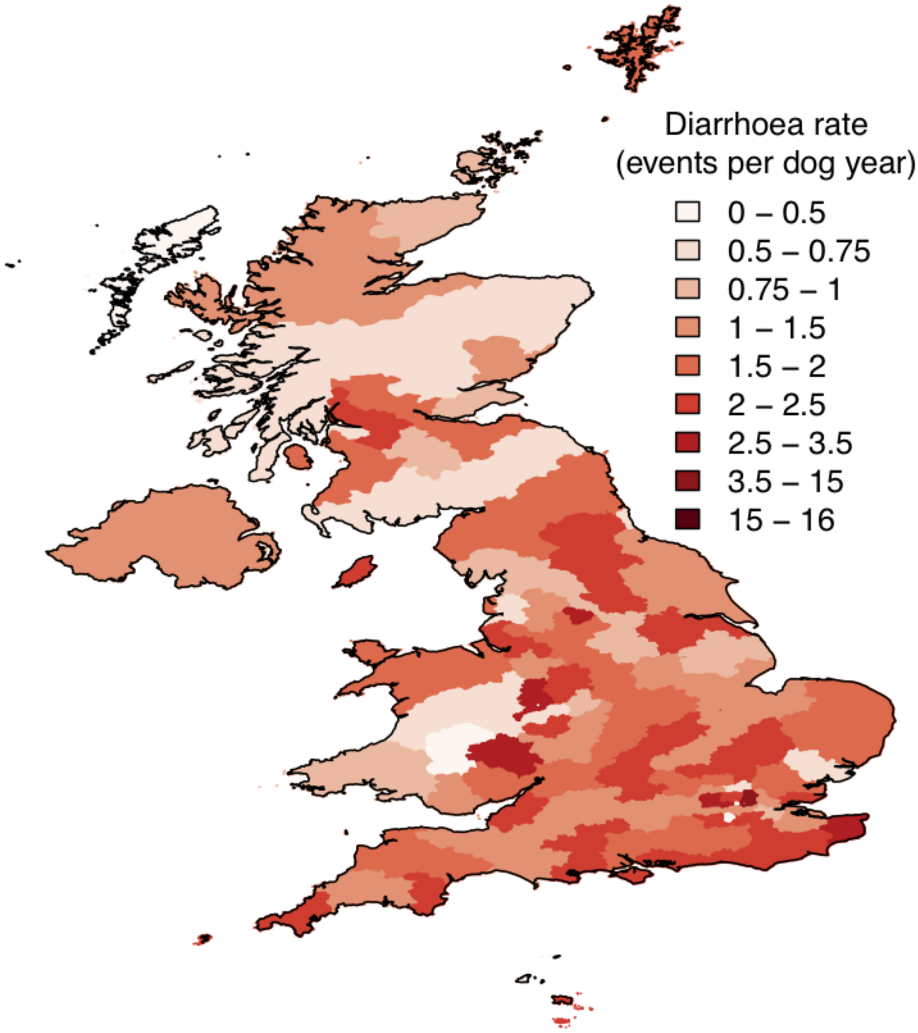


Figure 5.10: Postcode area diarrhoea incidence rates.
Windows of 40 days applied.

Table 5.6: Univariable Cox proportional hazards models of the time to first diarrhoea reports

Variable	Total time at risk				40 day window			
	Hazard ratio	95% CI:		*P* value	Hazard ratio	95% CI:		*P* value
	e^{β}	lower	upper		e^{β}	lower	upper	
Dog Purpose†								
pet	-	-	-	-	-	-	-	-
working	0.756	0.611	0.935	**0.010**	0.793	0.639	0.983	**0.034**
breeding or show	0.460	0.202	1.05	0.065	0.493	0.065	3.72	0.493
other	1.27	0.937	1.73	0.124	1.03	0.749	1.41	0.861
not reported	0.190	0.063	0.574	**0.003**	0.227	0.074	0.696	**0.009**
Household type								
more than one adult	-	-	-	-	-	-	-	-
family	0.847	0.755	0.950	**0.005**	0.881	0.784	0.990	**0.033**
retired	0.830	0.685	1.01	0.057	0.781	0.639	0.954	**0.015**
single adult	1.00	0.823	1.23	0.967	0.934	0.758	1.15	0.522
not reported	0.554	0.328	0.938	**0.0279**	0.594	0.351	1.01	0.052
Other Dog								
no	-	-	-	-	-	-	-	-
yes	0.767	0.667	0.859	<**0.001**	0.772	0.675	0.883	<**0.001**
Country‡								
England	-	-	-	-	-	-	-	-
Scotland	0.809	0.690	0.949	**0.009**	0.807	0.687	0.947	**0.009**
Wales	0.785	0.574	1.07	0.130	0.743	0.527	1.05	0.090
Northern Ireland	0.746	0.449	1.24	0.256	0.842	0.491	1.44	0.531
Latitude	0.945	0.917	0.974	<**0.001**	0.935	0.903	0.969	<**0.001**
Longitude	1.03	0.998	1.06	0.066	1.04	1.00	1.08	**0.039**
Density♣	1.01	1.01	1.01	<**0.001**	1.01	1.01	1.01	<**0.001**

†Assistance, multi-purpose and all other dogs were grouped as ‘other’.
‡Jersey, Guernsey and the Isle of Man were included in the analysis but excluded from reporting due to limited dog numbers.
♣ Density of people (per 100 people per km^2). Relates to England, Wales and Scotland only.

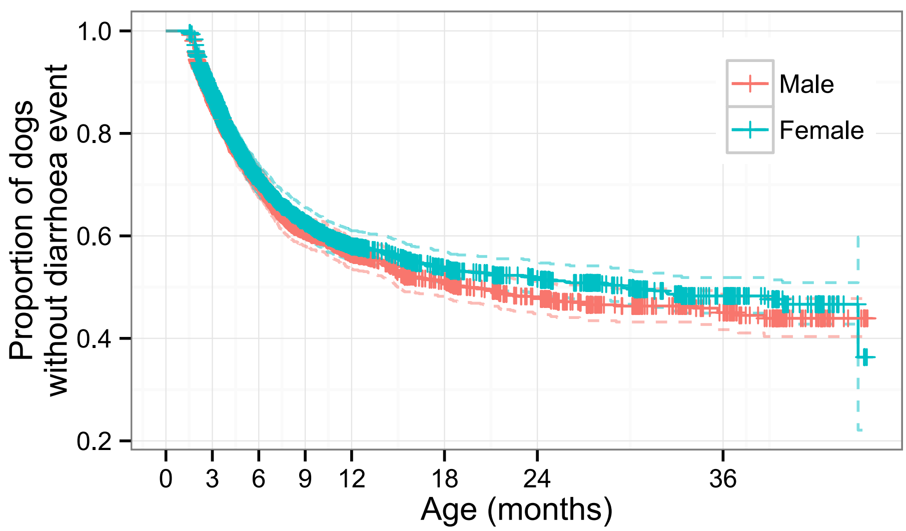


Figure 5.11: Time until report of first diarrhoea incident split by sex

time to first report of diarrhoea was longer for family households than those with more than one adult but if the window of 40 days was applied, retired households were also slower to report diarrhoea events in their dogs. Dogs whose owners did not report their household type or the dog's purpose were older at time of first diarrhoea report compared with those who reported these details, irrespective of the time at risk. This latter finding may relate to the likelihood of these owners making an illness report.

Owner smoking status, dog sex and coat colour and whether the household included a cat or any other pet (excluding dogs), were not found to be associated with time to reported diarrhoea. However, whilst the difference was not significant at the 5% level, male dogs seemed to succumb to diarrhoea earlier than female dogs (Figure 5.11). As diarrhoea incidents continue to accrue in the cohort, it will be interesting to see whether it will be possible to distinguish a difference between the two sexes.

When all potentially associated factors were combined into one model, windows of 40 days had to be applied to facilitate the inclusion of month of event. Final model choice was not simple because multiple models including different combinations of longitude, latitude and country explained similar levels of variation. For example, lower longitude was not associated with an overall decreased hazard of diarrhoea but living in Wales, which has a lower longitude, was associated with a reduced hazard. By contrast, latitude was positively associated with a lower hazard but Scotland, which has a higher latitude, was no longer associated with any hazard change. The final model explained 8.4% of a possible 38% of variation and is shown in Table 5.7. Seasonality and reduction of rate with age, captured by the individual months and their interaction with age, explained the majority of variation in the model. The interactions terms for all months were identical, indicating that the effect of seasonality decreased with age. Having another dog in the household, being in a family or retired household compared to a household with more than one adult and various geographic factors were associated with a reduced hazard of reporting diarrhoea. The model includes only dogs in Scotland, England and Wales because human population density was an important factor and these data were not available for NI or the various islands.

Table 5.7: Multivariable analysis of the time to first diarrhoea reports using a Cox proportional hazards model

Variable	Hazard ratio e^{β}	95% CI: lower	upper	*P* value
Other Dog				
no	-	-	-	-
yes	0.80	0.70	0.91	<**0.001**
Dog Purpose†				
pet	-	-	-	-
working	0.89	0.70	1.12	0.31
other	1.05	0.72	1.53	0.80
not reported	0.40	0.23	0.73	**0.002**
Household type				
more than one adult	-	-	-	-
family	0.837	0.742	0.944	**0.004**
retired	0.740	0.606	0.902	**0.003**
single adult	0.873	0.696	1.10	0.24
not reported	1.12	0.472	2.65	0.80
Latitude	0.935	0.887	0.985	**0.011**
Density‡	1.01	1.00	1.01	<**0.001**

Continued on next page

Table 5.7: Mulitvariable associations (continued)

Variable	Hazard ratio e^{β}	95% CI: lower	upper	P value
Country‡				
England	-	-	-	-
Scotland	1.13	0.88	1.44	0.35
Wales	0.71	0.52	0.98	**0.036**
Month				
Jan	1.10	0.71	1.69	0.67
Feb	1.31	0.79	2.18	0.29
Mar	1.50	0.95	2.38	0.081
Apr	1.20	0.75	1.94	0.44
May	-	-	-	-
Jun	1.08	0.69	1.69	0.73
Jul	1.01	0.67	1.52	0.97
Aug	1.81	1.23	2.65	**0.002**
Sep	1.93	1.32	2.83	<**0.001**
Oct	1.22	0.82	1.81	0.33
Nov	1.30	0.87	1.94	0.20
Dec	1.26	0.81	1.94	0.30
Month*Age				
All months	0.83	0.82	0.85	<**0.001**

†All non-pets and non-working dogs were grouped as ‘other’.

‡Density of people (per 100 people per km^2) was only available for England, Wales and Scotland.

5.3.4 Vomiting

Vomiting was reported 1,579 times; 25.7% (95% CI: 24.6 - 26.8%) of all illness reports. It was associated with 977 dogs which is 30.2% (95% CI: 28.7 - 31.9%) of all dogs that had reported data and 46.5% (95% 44.4 - 48.7%) of dogs that had illness reports. The reports were coded, as appropriate, with two separate signs: ‘Vomiting - other’ and ‘Vomiting - haematemesis’. The syndromes associated with ‘Vomiting - haematemesis’ are shown in Table 5.8 and those associated with ‘Vomiting - other’ are shown in Table 5.9.

Table 5.8: Other signs reported with 'Vomiting - haematemesis' and the syndrome frequency

Other sign	Other sign	Freq	Proportion involving vet visit
none	none	3	0.67
Faecal appearance abnormal - *diarrhoea*	none	3	0.67
Coughing	none	1	1.00
Dietary indiscretion - *other*	none	1	1.00
Faecal appearance abnormal - *diarrhoea*	Faecal appearance abnormal - *haematochezia*	1	1.00
Coughing	Vomiting - *other*	1	1.00

Table 5.9: Frequency of signs reported with 'Vomiting - other'

Other sign	Other sign	Freq	Proportion involving vet visit *95% CI*
none	none	1094	0.16 *(0.14 - 0.18)*
Faecal appearance abnormal - *diarrhoea*	none	363	0.47 *(0.42 - 0.53)*
Coughing	none	25	0.84 *(0.64 - 0.95)*
Pruritus	none	20	0.35 *(0.15 - 0.59)*
Vomiting - *haematemesis*	none	13	0.08 *(0.002 - 0.36)*
Dietary indiscretion - *other*	none	9	1.00
Dietary indiscretion - *foreign body ingestion*	none	6	0.67

Continued on next page

Table 5.9: 'Vomiting - other' syndromes (continued)

Other sign	Other sign	Freq	Proportion involving vet visit
Faecal appearance abnormal - *diarrhoea*	Vomiting - *haematemesis*	5	0.40
Dietary indiscretion - *other*	Faecal appearance abnormal - *diarrhoea*	3	1.00
Mass/swelling - *oral (mouth)*	none	3	1.00
Seizure(s)	none	3	1.00
Coughing	Gagging/retching	2	1.00
Dietary indiscretion - *foreign body ingestion*	Faecal appearance abnormal - *diarrhoea*	2	1.00
Wound	none	2	1.00
Gait abnormality - *lameness*	Pruritus	1	1.00
Faecal appearance abnormal - *diarrhoea*	Pyrexia/hyperthermia	1	1.00
Lethargy	none	1	1.00
Faecal appearance abnormal - *diarrhoea*	Lymphadenomegaly	1	1.00
Coughing	Dietary indiscretion - *foreign body ingestion*	1	1.00
Faecal appearance abnormal - *diarrhoea*	Faecal appearance abnormal - *haematochezia*	1	1.00
Skin (cutaneous) abnormality - *eruptions/hives/rash*	none	1	1.00
Pyrexia/hyperthermia	none	1	1.00
Gait abnormality - *lameness*	none	1	1.00

Continued on next page

Table 5.9: 'Vomiting - other' syndromes (continued)

Other sign	Other sign	Freq	Proportion involving vet visit
Mass/swelling - *other*	Pruritus	1	1.00
Appetite decreased	Lethargy	1	0.00
Coughing	Vomiting - *haematemesis*	1	1.00
Skin (cutaneous) abnormality - *other*	none	1	1.00
Coughing	Presenting complaint not listed	1	0.00
Presenting complaint not listed	Pruritus	1	1.00

One syndrome included four signs: 'Vomiting - *other*', 'Coughing', 'Discharge - *ocular (eye)*' and 'Opthalmic (eye) abnormality'.

Both frequency and proportion involving vet visit were one.

As single signs 'Vomiting - other' and 'Vomiting - haematemesis' were reported 1,094 and three times respectively. Focussing just on the 'Vomiting - other' reports, 1,030 had associated start and end dates and, of these, 833 reports (80.9%, 95% CI: 78.3 - 83.2%) indicated that the vomiting lasted less than one day. In terms of frequency, 656 reports indicated that the dog had vomited just once but oddly 26 of these apparently lasted longer than one day. Excluding the episodes that involved one incident of vomiting because they entirely dominated the plot, the distribution of duration and frequency is shown in Figure 5.12.

The proportion of each of these groups of signs, and those that were reported to have only happened once, that involved a veterinary visit are shown in Figure 5.13. As might be expected, the trend is that the likelihood of a veterinary visit increased as the duration and frequency of the vomiting increased.

The numbers of vomiting reports were reduced from 1,579 to 1,541 when fitted into the total time at risk; this implies that 38 reports took place more than 28 days before the first data entry. The overall incidence rate for the cohort was the

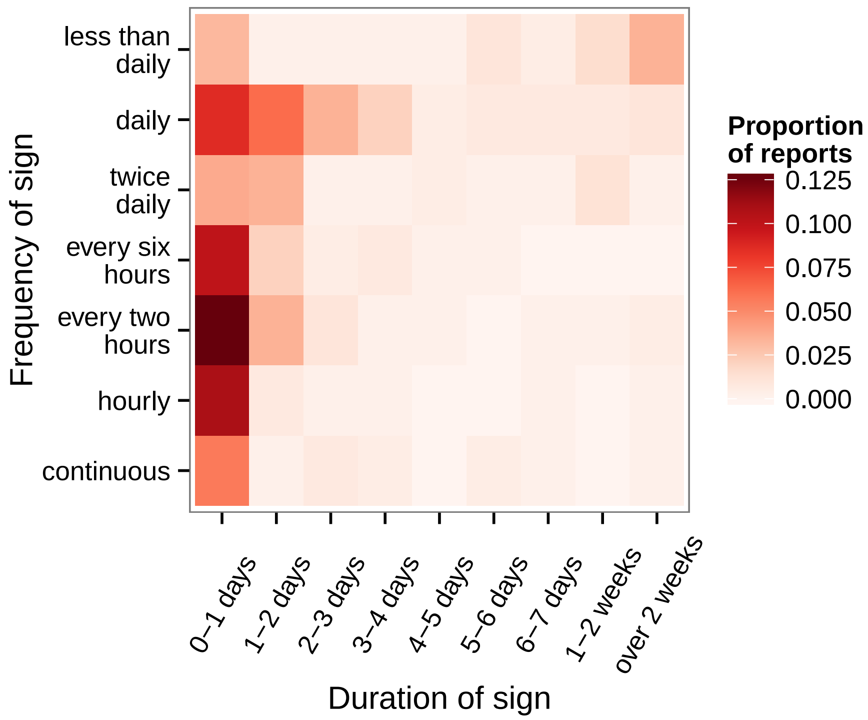


Figure 5.12: Distribution of vomiting reports according to duration and frequency of sign (all with frequency of 'once' omitted)

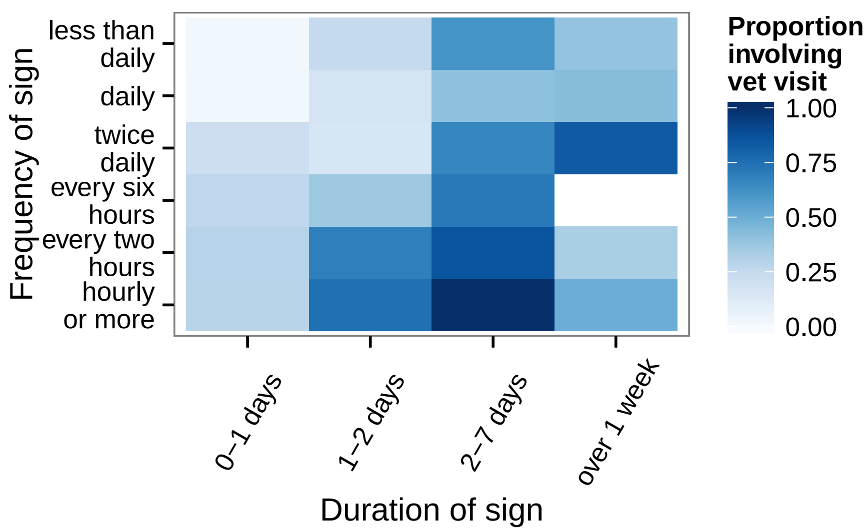


Figure 5.13: Proportion of vomiting reports that involved a veterinary visit.

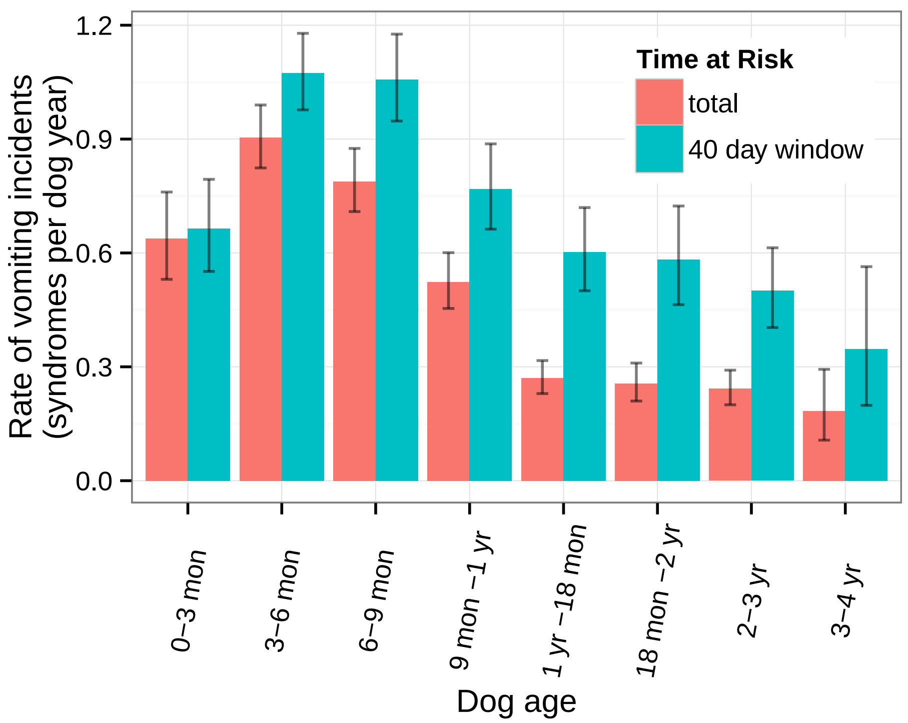


Figure 5.14: Age specific vomiting rates. Total time at risk includes 1,541 incidents in 3,098 years at risk but this drops to 1,404 incidents in 1,733 years when using 40 day windows. Bars represent 95% poisson CI.

same as for diarrhoea, 0.50 (95% CI: 0.48 - 0.52) reports per dog year and, as can be seen in Figure 5.14, the rates depended heavily on the age of the dog.

Attempting to minimise the effect of recall decay by applying time at risk windows of 40 days prior to each data entry excluded a further 137 reports, reducing the numbers to 1,404 incidences and 1,733 dog years at risk. The overall incidence rate increased to 0.81 (95% CI: 0.77 - 0.85) reports per dog year. The highest rates were when the dogs were between three and nine months of age (Figure 5.14).

The maximum number of reports of vomiting for one dog was ten times which was reported for just one dog. The remaining frequencies of vomiting reports are included in Table 5.3. The majority of dogs that had vomiting reports had just one report (637 of 997 dogs). Using the total time at risk the maximum reported time at risk was 42 months for dogs that did and did not have reports of vomiting. This similarity was repeated when windows of 40 days were applied;

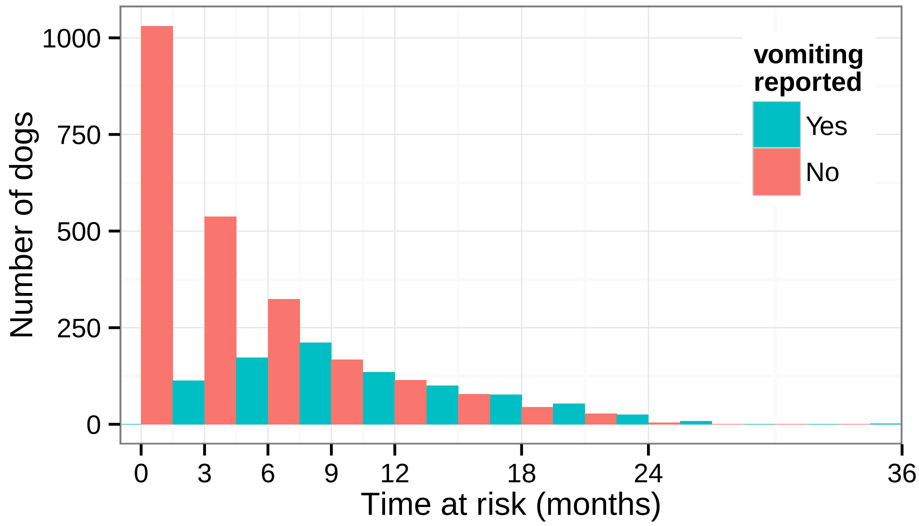


Figure 5.15: Time at risk for dogs that do and do not have reports of vomiting. Windows of 40 days applied. Dogs with zero time at risk excluded.

the maximum times at risk were 34 and 33 months for those that did and did not have reports respectively. However the distributions of times at risk were markedly different (Figure 5.15). A preponderance of dogs that did not have reports of vomiting contributed less time to the project.

The monthly incidence rates for vomiting (using the incidents fitted into the 40-day windows) were plotted (Figure 5.16). The confidence intervals largely overlapped but it appeared that January had a high rate. In order to account for the seasonal pattern of births in the cohort, the monthly rates were stratified according to the age of the dog when the syndrome first started (Figure 5.17). January continued to stand out, particularly for dogs under one year of age. A Cox proportional hazards model indicated that the hazards in January and September were 1.34 (95% CI: 1.03 - 1.74) and 1.30 (95% CI: 1.00 - 1.68) times that of the hazard in May respectively (Table 5.10). This reflects a similarly low rate in May and high rate in September for diarrhoea incidents but vomiting did not have the same peak in August that was found for diarrhoea and diarrhoea did not peak in January.

Unlike diarrhoea, vomiting incidence rates did not seem to have a geographical pattern (Figures 5.18 and 5.19).

The univariable Cox analyses (Table 5.11) confirmed the general impression

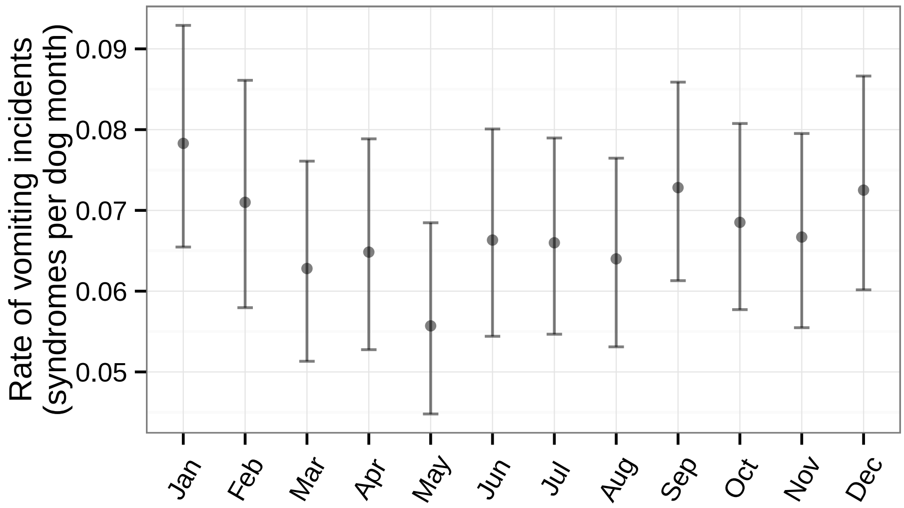


Figure 5.16: Monthly vomiting rates using incidents within windows of 40 days prior to each data entry.

given by Figures 5.18 and 5.19; the rate of vomiting did not appear to vary according to country or time to first reported event. Neither latitude nor longitude were associated with time to first vomiting event but there was a small association between increasing human population density and reduced time to first vomiting event. As with diarrhoea, having another dog in the household had an apparent protective effect. This protection was repeated when there was a cat in the household. Interestingly, associations did not typically vary when different times at risk were applied. In both cases, the time to first report of vomiting was longer for retired households than all other types and longer for working dogs than pets.

Owner smoking status, sex of the dog and coat colour and whether the household included any other pet (excluding dogs and cats), were not found to be associated with time to reported vomiting.

Combining the various factors into a multivariable model was complex. The final model shown in Table 5.12 explained just 0.8% of 56% of possible variation. This could be greatly improved to approximately 8% of 37% if month of occurrence was included but the hazards were not proportional. Efforts to deal with this resulted in models that failed to converge. There were not enough data to support such a complex approach.

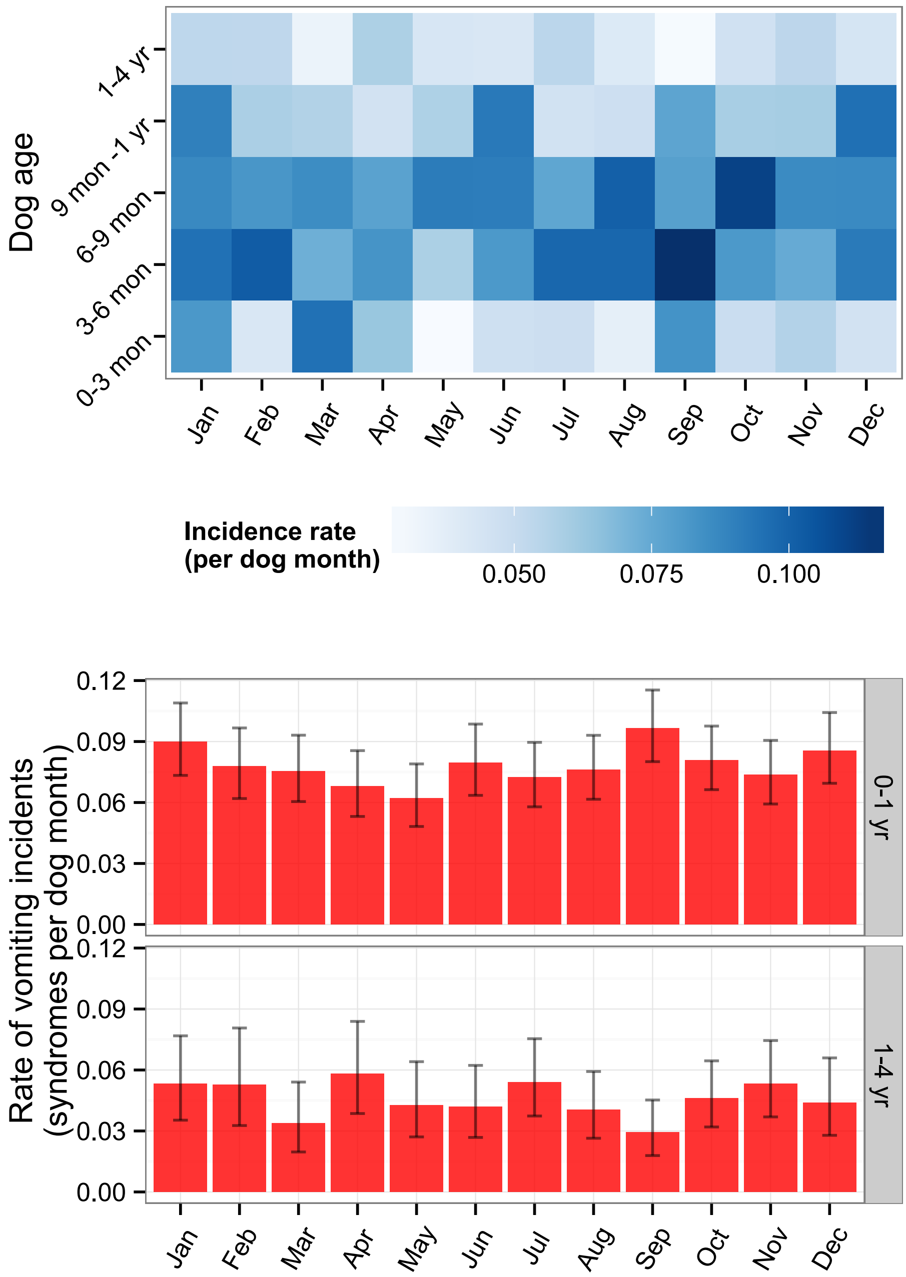


Figure 5.17: Age specific monthly vomiting rates using incidents within windows of 40 days prior to each data entry.

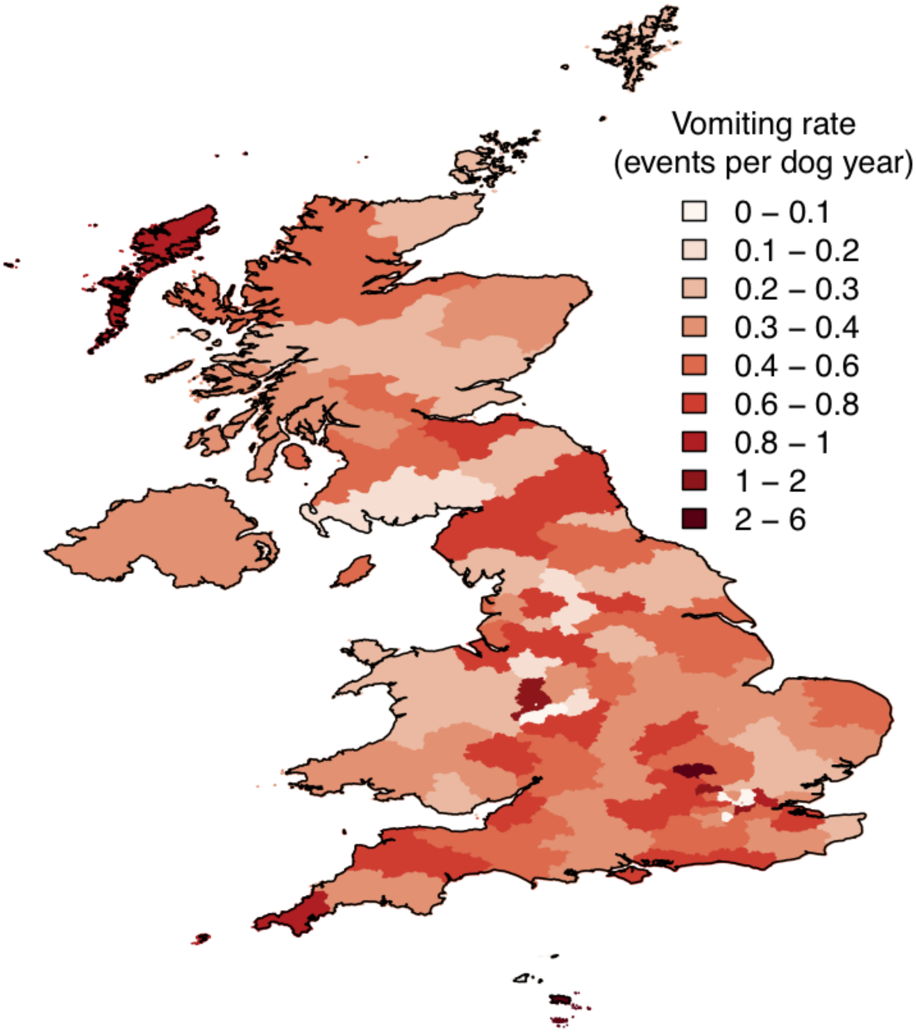


Figure 5.18: Postcode area vomiting incidence rates.
Total time at risk.

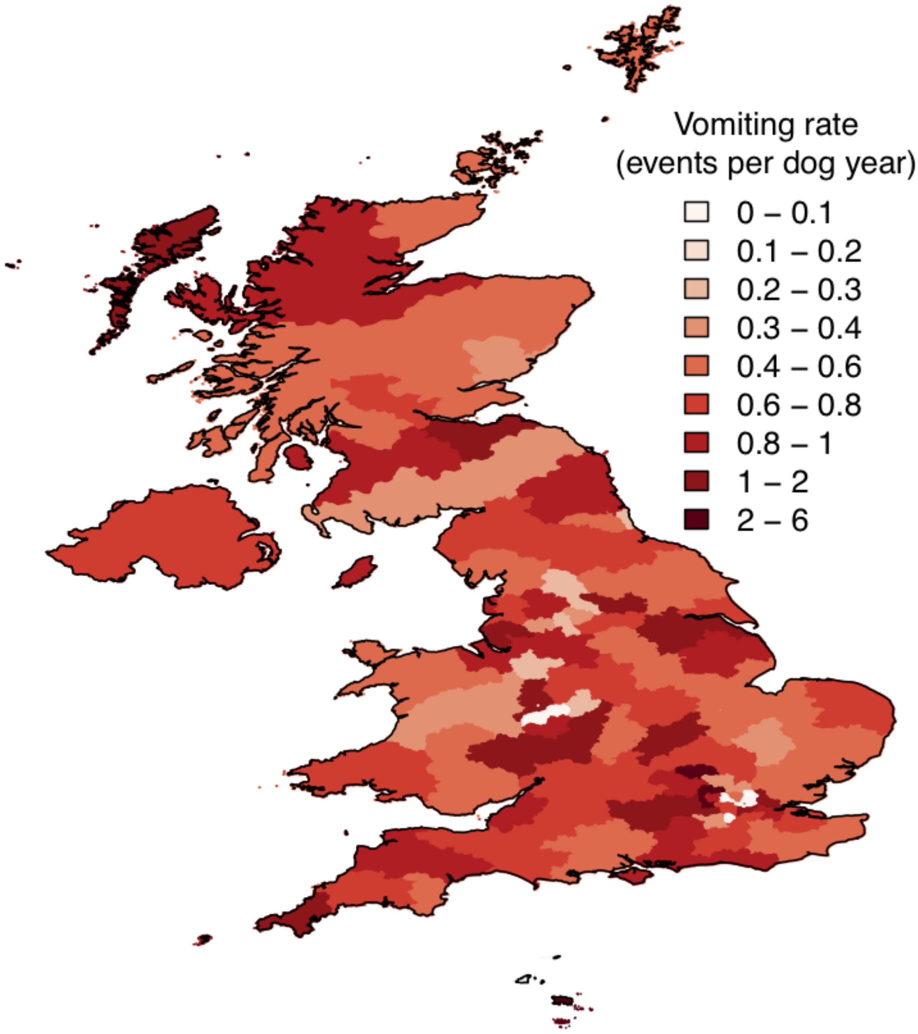


Figure 5.19: Postcode area vomiting incidence rates.
Windows of 40 days applied.

Table 5.10: Cox proportional hazards model of association between month of the year and time to first vomiting event

Month	Hazard Ratio e^{β}	95% CI: lower	upper	P-value
January	1.34	1.03	1.74	**0.03**
February	1.21	0.921	1.59	0.17
March	1.11	0.840	1.46	0.47
April	1.15	0.866	1.52	0.34
May	1	-	-	-
June	1.20	0.916	1.58	0.19
July	1.19	0.913	1.56	0.20
August	1.15	0.884	1.50	0.30
September	1.30	1.00	1.68	**0.05**
October	1.20	0.930	1.54	0.16
November	1.15	0.878	1.51	0.31
December	1.23	0.942	1.60	0.13

Table 5.11: Univariable Cox proportional hazards models of the time to vomiting reports

Variable	Total time at risk				40 day windows			
	Hazard ratio	95% CI:		*P* value	Hazard ratio	95% CI:		*P* value
	e^{β}	lower	upper		e^{β}	lower	upper	
Dog Purpose†								
pet	-	-	-	-	-	-	-	-
working	0.443	0.325	0.605	<**0.001**	0.442	0.319	0.611	<**0.001**
other	0.691	0.458	1.04	0.077	0.797	0.534	1.19	0.267
not reported	2.79e-07	2.10e-07	3.70e-07	<**0.001**	2.81e-07	2.15e-07	3.69e-07	<**0.001**
Household type								
retired	-	-	-	-	-	-	-	-
family	1.49	1.18	1.89	<**0.001**	1.72	1.33	2.21	<**0.001**
more than one adult	1.52	1.20	1.93	<**0.001**	1.72	1.34	2.22	<**0.001**
single adult	1.39	1.01	1.91	**0.042**	1.44	1.02	2.02	**0.039**
not reported	1.45	0.906	2.32	0.122	1.65	1.00	2.72	**0.049**
Other Dog								
no	-	-	-	-	-	-	-	-
yes	0.656	0.578	0.745	<**0.001**	0.659	0.579	0.750	<**0.001**
Cat								
no	-	-	-	-	-	-	-	-
yes	0.849	0.738	0.978	**0.023**	0.855	0.740	0.989	**0.034**
Country‡								
England	-	-	-	-	-	-	-	-
Scotland	0.957	0.814	1.13	0.597	0.956	0.811	0.1.13	0.593
Wales	0.774	0.532	1.13	0.182	0.809	0.558	1.17	0.263
Northern Ireland	0.926	0.560	1.53	0.765	0.964	0.561	1.66	0.895
Latitude	0.985	0.954	1.02	0.358	0.979	0.944	1.02	0.252
Longitude	0.983	0.950	1.02	0.326	0.984	0.945	1.03	0.451
Density♣	1.01	1.00	1.01	**0.01**	1.01	1.00	1.01	**0.006**

†Assistance, multi-purpose and all other dogs were grouped as ‘other’.
‡Jersey, Guernsey and the Isle of Mann were included in the analysis but excluded from reporting due to limited dog numbers.
♣ Density of people (per 100 people per km^2). Relates to England, Wales and Scotland only.

Human population density did not improve the fit of the final, relatively simple, multivariable model. Despite there being no obvious association between latitude, longitude and vomiting rates (Figures 5.18 and 5.19), including both latitude and longitude improved the model. In particular, increasing latitude was associated with longer time to first vomiting event. Having another dog or a cat in the household remained strongly protective in the combined model.

Table 5.12: Multivariable analysis of the time to first vomiting reports using a Cox proportional hazards model

Variable	Hazard ratio e^{β}	95% CI: lower	upper	P-value
Other Dog				
no	-	-	-	-
yes	0.69	0.59	0.80	<**0.001**
Cat				
no	-	-	-	-
yes	0.80	0.68	0.94	**0.006**
Dog Purpose†				
pet	-	-	-	-
working	0.47	0.33	0.67	<**0.001**
other	0.88	0.56	1.38	0.58
not reported	2.6e-07	2.0e-07	3.5e-07	<**0.001**
Household type				
retired	-	-	-	-
more than one adult	1.85	1.40	2.45	<**0.001**
family	1.81	1.37	2.41	<**0.001**
single adult	1.56	1.06	2.28	**0.02**
not reported	1.46	0.60	3.52	0.40
Latitude	0.96	0.92	0.998	**0.038**
Longitude	0.98	0.93	1.02	0.28

†All non-pets and non-working dogs were grouped as 'other'.

5.4 Discussion

There were 6,115 signs of illness reported to Dogslife in the first three and a half years of the project. They ranged from thousands of instances of diarrhoea and vomiting to a single report of a cataract. Approximately half of the signs were not associated with a veterinary visit. The validation process (Section 3.4.4) indicated that under-reporting of veterinary visiting illness instances was 22% so 6,115 should be regarded as an underestimate. As a general trend, the illness rates decreased as the dogs aged which reflects similar findings from insured, vet-visiting dogs in Sweden (Egenvall et al. [2000]).

In contrast to a recent study of 3,884 vet-visiting dogs that found otitis externa to be the most prevalent condition (O'Neill et al. [2014b]), ear-related complaints were only the sixth most numerous of the twenty event types most frequently reported to Dogslife. This might simply be due to breed susceptibility or the younger age of the Dogslife cohort, which was under four years, compared to a median age of 4.8 years for a variety of breeds in the O'Neill et al. study. However it also seems likely that there is a difference between illnesses that do, and do not, precipitate vet visits.

Of the twenty most frequently reported syndromes, the proportion associated with a veterinary visit ranged between 16% for vomiting and 88% for mass/swelling and anal irritation. There is a similar phenomenon which was initially described in human medicine in 1963 called the *Symptom Iceberg* (Last and Adelaide [1963]). The authors detailed the signs detected by General Practitioners (GPs) in England and Wales and estimated undetected levels of morbidity from population morbidity and mortality statistics. They suggested that the signs seen by GPs could be thought of as the portion of the iceberg seen above the water with unreported signs of illness hidden under the water. The authors of the 1963 study found, much like Dogslife, that the proportions of signs estimated to go unseen by GPs was heavily dependent on the type of the sign.

Since 1963, considerable effort has been made to understand what drives people to seek medical attention and to try to inform the public regarding which signs they should see a doctor about. To quote Elnegaard et al. [2015], from a public health perspective, "symptoms potentially indicative of a serious disease should *preferebly* lead to healthcare seeking, while other symptoms should not". The implication of the iceberg analogy is that the proportion of signs that are not seen by medical practitioners is far higher than those seen by practitioners. This is supported by multiple human studies such as that by Dahlquist et al. [1984], which found that just 10% of signs involved seeking medical attention. The study by Elnegaard et al. [2015] was based online and involved 49,706 Danish subjects. They asked participants about specific symptoms during the preceeding four weeks and found that 37% of participants visited their doctor with at least one sign of illness. Whilst not strictly the same proportion as it considers individuals rather than signs, 37% is closer to Dogslife's 47% of signs involving a veterinary visit than the 10% of Dahlquist et al. [1984] but the authors suggested 37% was an unusually high proportion.

There is a clear distinction between canine and human medicine. In canine medicine, we must rely on owners noticing a sign of illness before they can report it and if it produces minimal behavioural change, the owner is unlikely to do so. The study by Elnegaard et al. [2015] included questions about tiredness and headaches and found that that 20.2% and 17.7% of reports respectively were associated with medical visits. Both of these are relatively low percentages, reducing the overall average and neither sign would likely be reported by a Dogslife owner because they might not produce noticeable behavioural changes in dogs. Whilst the proportion of signs involving a veterinary visit according to Dogslife is high in human terms, such direct comparisons may not be appropriate. Instead, Dogslife data provide the first estimates of population disease burden considering signs that do and do not precipitate veterinary visits. They might be compared with smaller studies that look at specific illnesses and hopefully, like the study by Last and Adelaide [1963], provide a benchmark for future studies of the overall *Symptom Iceberg* in dogs.

It might be assumed that owner perception of illness severity governs the likelihood of an owner taking their dog to a vet or perhaps a perception of whether the sign is likely to resolve itself without intervention. By extension, it appears that Dogslife owners did not perceive diarrhoea, vomiting and pruritus to be particularly serious. Vomiting in particular had a very low level of veterinary visitation, perhaps because it was predominantly reported to happen just once per report.

For the first time, the burden of non-veterinary visiting illnesses on young LR has been quantified for thousands of animals. Three of the five most frequently reported syndromes were gastrointestinal, comprising vomiting and/or diarrhoea and typically not associated with a vet visit. This is a large burden, particularly in young, growing dogs, so interventions that produce small percentage reductions in incidents would impact a large number of dogs and owners. Vomiting and diarrhoea might largely be considered unpleasant rather than dangerous but reducing incidence, particularly for dogs that suffered repeated events, would have a positive impact on the welfare of dogs and their owners. A study in the USA found that 18.5% of 1,984 owners reported soiling as a reason for relinquishing their dog to a rescue shelter (Salman et al. [2000]). There is a qualitative difference between stress-based soiling and illness-based soiling but, from an owner's perspective, the short-term impact is similar.

5.4.1 Recall Decay

Windows of 40 days prior to each data entry were applied to try to minimise recall decay but this might have introduced a different bias. Owners were able to answer the questionnaire at any time and were actively encouraged to make a data entry if their dog became sick. Whilst cleaning and analysing the data, it became clear that many owners answered the questionnaire on the day of a veterinary visit or illness incident. If the behaviour of these owners was consistent and they always came to the website when their dog was ill, then the time disregarded for these dogs was time when the dogs were genuinely healthy. Applying the windows would have inflated illness rates inappropriately. Nevertheless, the validation exercise indicated that illness reporting was subject to recall decay and failing to try to account for it would be equally inappropriate. By using both the total time at risk and the time at risk generated by applying 40 day windows, two different pictures were created of illness rates and times to event. Where they agreed, estimates generated are likely to be reliable but where they were different, which was more correct would be debatable and the reasons for differences can only be speculated upon.

Rates of vomiting and diarrhoea were determined using both the total time at risk and 40 day windows. Applying windows would reduce the power of models to detect associations because it reduced the number of illnesses under consideration. Conversely it might also increase the power by more narrowly defining times when the dogs were definitely not ill. Aspects of the analysis that were affected by recall decay would be highlighted by comparing the analyses of the different times at risk. The idea that there might be differential recall bias was supported somewhat by the univariable Cox analysis of time to diarrhoea and household type. Dogs from retired households were more likely to be associated with a delay to reporting of diarrhoea once the windows of 40 days were applied. These windows reduce incidents and time under consideration and for retired households, they presumably disproportionately reduced the time at risk when compared to other household types. Longitude was the only other variable whereby the confidence interval shifted further from one when windows were applied to diarrhoea incidents.

The pattern of households being disproportionately affected by recall decay was repeated with vomiting. When 40 day windows were applied, the hazard ratios increased for all household types in comparison to retired households. In all cases, dogs from retired households had the longest delay to reports of vomiting. In the univariable analyses, neither country, latitude nor longitude were associated with time to vomiting, regardless of which time at risk was considered.

5.4.2 Comparison of Vomiting and Diarrhoea

In the first three and a half years of Dogslife, both vomiting and diarrhoea were reported over 1,000 times. Additionally, they were reported 363 times together. Such large numbers of incidents allowed for in depth analysis and comparison of the two signs. Only 31% of diarrhoea reports were associated with a veterinary visit but this was nearly double that of vomiting at just 16%. This was considerably higher than the values found in a study of 772 dogs in the UK in 2007 that found that 10% of dogs that had diarrhoea and 5% of dogs with vomiting were taken to the vet (Hubbard et al. [2007]). The Hubbard et al. study asked participants about the two weeks following receipt of the questionnaire so there is a suggestion that different methodology might be affecting results. In particular, recall decay in Dogslife might differentially be affecting illness reports; incidents with perceived lower severity (such as those that do not precipitate a veterinary visit) have been found to be more likely to be forgotten compared to more serious illness events (US Department of Health, Education and Welfare [1961]) so perhaps owners were forgetting to report the non-vet-visiting incidents of diarrhoea and vomiting to Dogslife.

There was presumably greater concern on behalf of owners when the two signs occurred together because, when they were combined, owners took their dog to the vet in relation to 47% of events. This phenomenon has also been described in human health with a positive linear relationship between the concurrent number of possible signs and likelihood of visiting a GP (Elnegaard et al. [2015]). In Dogslife, duration of sign rather than frequency appeared to be the primary motivation for veterinary visits when the dog had diarrhoea but there was more of an emphasis on frequency for vomiting. The Hubbard et al. [2007] study found that if a diarrhoea incident lasted for two of more days, 66% of reports involved a veterinary visit but they did not have any reports of vomiting that lasted more than two days. Interestingly, diarrhoea reports to Dogslife predominantly needed to last at least two days before the owner would take their dog to the vet but, for vomiting reports that were happening at least every two hours, the owner would take their dog to the vet after just one day. Despite vomiting precipitating a lower proportion of veterinary visits than diarrhoea, if vomiting was frequent enough, owners would apparently react more quickly.

Both vomiting and diarrhoea had incidence rates that peaked between three and six months of age. Many factors would likely reduce the number of incidents in dog under three months of age. They would be partially protected by maternal immunity and initially consume a diet of just milk. Before completing early courses of vaccinations at approximately 12 weeks of age, owners are advised not to expose their puppy to other animals (Kennel Club [2015]) and, as such, the dogs would be exposed to fewer infectious agents. As they aged, it might be hypothesised that these young dogs would simultaneously lose maternal immunity and be exposed to a gamut of gastrointestinal challenges as they explored their surroundings. Relative youth also appears to be associated with a seasonal pattern of reporting. For dogs under one year, the lowest incidence rates for both vomiting and diarrhoea occurred in May and there were peaks in September for both signs. However there were also distinct differences, including a high incidence rate for vomiting in January. This was perhaps associated with consumption, and subsequent rejection, of holiday food. The seasonal pattern of diarrhoea and vomiting incidence for Dogslife does not agree with findings from a study of four large breeds in Norway (Sævik et al. [2012]). The Norwegian study found peaks in Summer (June - August) for both signs at all ages and suggests that the peaks (with the exception of vomiting in January) may be driven by climactic factors which would clearly differ between the two studies.

Multiple studies have found higher rates of diarrhoea in males compared to female dogs (Hubbard et al. [2007]; Stavisky et al. [2011]; Sævik et al. [2012]). It is interesting that despite having greater numbers of dogs than all of these studies (two of which were also based in the UK), there was no significant difference found between male and female Dogslife dogs. The analysis was based on 3,234 dogs until 31st December 2013 and male dogs appeared to succumb sooner than females but not significantly at the 5% level. In August 2015, Dogslife had over 5,500 registered members and ideally the analysis would be repeated with these extra dogs. Such an analysis might give the power required to find a difference between the sexes or definitively indicate that the Dogslife population was different to those studied above. It should be noted that all of the other studies had a wider range of ages and breeds in their studied populations so it is possible that young Dogslife LRs simply do not have a sex difference in terms of diarrhoea risk.

Difficulties with convergence of the multivariable vomiting model makes comparisons between these models for vomiting and diarrhoea problematic. Direct comparison between the univariable analyses remains feasible and presents an interesting picture. Time to the first reports of diarrhoea and vomiting was lower in household pets than working dogs and longer in retired households than any other household type. The increased time to event for working dogs was more exaggerated for vomiting than diarrhoea. As with all Dogslife data, it is unclear whether there was a difference between the groups of dogs, a difference in the owners' abilities to detect a sign, or a difference in the likelihood of reporting. Nevertheless, there appeared to be a strong geographical pattern to diarrhoea reporting that was not repeated for vomiting. Time to first diarrhoea report was lower at lower latitudes and higher longitudes and significantly positively correlated with human population density. For vomiting, there was a univariable correlation between increasing human population density and decreased time to first vomiting incident but this was not a significant factor in the multivariable model. Time to first vomiting report was not univariably associated with country location, nor latitude and longitude but latitude was significantly associated in the multivariable model. The initial impression given by Figures 5.18 and 5.19 that there was no geographical pattern underlying vomiting was not supported by the numerical analyses but the pattern was not as clear as for diarrhoea. One might hypothesise that more of the diarrhoea reporting than the vomiting had an infectious aetiology. Greater human population density would, on average, be associated with a greater number of potential infectious contacts.

The Norwegian study of four large breeds (Sævik et al. [2012]) previously demonstrated that dogs in urban areas were at greater risk of developing diarrhoea than those in suburban or rural areas but they found no association for vomiting. Unfortunately Scotland uses a different method for describing urban/rural distinctions than NI and they are both different to the system used by England and Wales so assessing urban and rural risks was impractical with Dogslife data. Nevertheless, urban/rural classifications are typically based on human population density and the findings associating diarrhoea with higher levels of human population density seem to support the findings of the Norwegian study. The vomiting results were less clear-cut and they do not agree with the Norwegian study which found no effect. As with the work assessing risk according to the sex of the dog, it would be interesting to repeat these analyses with larger numbers of animals to determine whether greater power would make it possible to refine the results of the risk factor analyses for vomiting and geographical variation.

The final differences between vomiting and diarrhoea were with regard to other pets being in the household. Having a cat in the household was associated with a reduced hazard of vomiting reporting and having another dog was associated with a smaller hazard for vomiting than diarrhoea. On might imagine that having other pets would increase exposure to infectious disease so it is surprising that having another dog or a cat were associated with lower hazard. It is difficult to know whether these associations are genuine, perhaps because of more robust immune function, or whether the owners of these dogs are simply less likely to notice or report diarrhoea and vomiting. Ideally this analysis would be repeated in other studies to help understand whether it is a true reflection of decreased risk in multi-animal households.

Despite often being reported together, vomiting and diarrhoea were associated with different risk factors suggesting that the two signs have different aetiologies. On that basis, future investigations of gastrointestinal disease should treat vomiting and diarrhoea as different processes or risk missing detail.

5.4.3 Conclusion

As detailed in Appendix 6, the Dogslife illness records are extensive. The in-depth analysis of vomiting and diarrhoea demonstrate the utility of these data for detailed analyses. The risk factor analyses have built on previous studies and added novel information regarding household type and dog purpose, the risks of owning other pets and geographical patterns of illness across the UK. For rarer signs, because of the large numbers of participants, appreciable numbers of cases are being accrued that might be used in targeted case-control studies. The next step for Dogslife will be to assess the DNA samples collected and continue linking these illness reports to the wealth of demographic and lifestyle data in order to develop an understanding of the associations between canine lifestyle, genetics and health.

Chapter 6

Exercise Validation

6.1 Introduction

In exercise science, the gold standard measure of physical activity is calorimetry (Melanson and Freedson [1996]) whereby total energy expenditure is calculated by administering heavy water and measuring the levels of heavy isotopes remaining in urine (Coward [1988]). This is impractical outside laboratory settings and does not assess the energy expended whilst undertaking different activities, for example, comparisons of resting, walking and running. More versatile tools such as questionnaires (Innes and Barr [1998]), Global Positioning System (GPS) collars (Bruno et al. [2015]) and accelerometers (Dow et al. [2009]) have been widely used to assess activity in less controlled settings.

Each method has advantages and disadvantages, for example, questionnaires do not necessitate the purchase and management of equipment so they are ideal for large numbers of participants. Unfortunately, as discussed previously, using questionnaires introduces issues of validity. GPS collars allow for distance travelled in a period of time to be assessed. With refinement, the data may also be manipulated to give each dog's velocity and acceleration. Unfortunately, whilst they have been used to distinguish between healthy and osteoarthritic dogs in an epidemiological study of 17 dogs (Bruno et al. [2015]), GPS collars may only be used outside so they cannot be used to assess total daily activity. They currently also have high power requirements making them impractical for long-term use by untrained people. Accelerometers are not required to repeatedly contact satellites so they have longer battery lives than GPS collars and can be used indoors, however they still require a financial outlay and cannot practically be used in large-scale studies.

The scale of the Dogslife cohort dictates that questionnaire measurement was the only practical way to measure activity. As discussed in Section 4.2.9, the exercise section of the Dogslife questionnaire (Appendix 2, Section A2.4) was designed to try to capture the time spent on a range of different activities. It was hoped that answers to the questionnaire would serve as a proxy for each dog's energy expenditure and more, that it would also facilitate investigations of specific activities which may be related to musculoskeletal problems (for example Krontveit et al. [2012] found an association between stair climbing in dogs under three months of age and subsequent development of hip dysplasia). Ideally the data collected about specific activities would be appropriate for use in risk factor analyses such as those undertaken in Section 4.3.10. It was shown that univariably, owner reports of total daily time spent exercising, time spent fetching, chasing and retrieving and time spent on other activities were all negatively correlated with dog weight and time spent walking on the lead was positively correlated with dog weight.

Ideally, each completed set of exercise questions would also be a good proxy for that dog's activity, similar to results generated by the use of an accelerometer. This chapter describes efforts to validate the exercise section of the questionnaire using accelerometer data collected from a subset of the cohort. The aim was to determine whether the exercise section of the Dogslife questionnaire might be more broadly used as a valid tool to assess activity in LR.

6.2 Methods

Accelerometers were sent to a subset of the cohort with a stamped addressed envelope to facilitate their return. Also included, and detailed in Appendix 7, were an A4 cover letter (Figure A12), an A5 flyer describing the study protocol (Figure A13) and a paper questionnaire (Figures A14 and A15). The owners were asked to email Dogslife when they received the accelerometer and email again with the times when they attached and detached the device from their dog's collar. They were asked to put the accelerometer on their dog's collar for a week and complete the questionnaire regarding their dog's activity for that week.

One week was chosen in accordance with findings reported by Dow et al. [2009] which emphasised the need to include both week days and weekend days. The paper questionnaire (Figures A14 and A15) comprised just two sides of A4 largely copied from the exercise section of the online questionnaire which is available in Appendix 2, Section A2.4. In addition there were two questions regarding how much time the dog spent in the car and whether the questions captured all of the dog's exercise. Upon return, the accelerometer data were downloaded and the answers from the paper questionnaire were entered into a spreadsheet by the project administrator.

6.2.1 Sampling Strategy

Two main premises guided which owners were asked to use the accelerometers. Following the assessments of exercise in Section 4.3.11, it was apparent that few dogs ran on the lead with their owners and working dogs exercised differently from household pets. In order to validate the questionnaire, sufficient dogs from each of these groups would need to be selected. In the absence of estimates of effect size, sample size power calculations were redundant. Instead it was decided to disproportionately sample from dogs that ran on the lead with their owner and working dogs such that approximately 10 of each category would be chosen in the first 100 sampled dogs. The chances of choosing a working dog were therefore multiplied by 1.5 and the chances of choosing a dog that ran on the lead were multiplied by 20. More generally, dogs were randomly selected from those that had a data entry in the previous month, were over one year of age and whose owners had given permission for Dogslife to contact them by email and/or telephone. Up to 20 dogs were chosen every month and the project administrator worked through the list sending accelerometers to owners who agreed to participate. Sampling began in June 2013 with three accelerometers. Recruitment was a dynamic process so the dogs were not necessarily part of the 4,307 dogs previously described in this thesis.

6.2.2 Accelerometers

Philips Respironics Actical accelerometers were used (Figure 6.1). The device specifications are given in Appendix 7, Figure A16. Each contains a piezoelectric

Figure 6.1: Philips Respironics Actical with 20 pence piece for scale.

sensor which generates a voltage in response to changes in velocity with time (acceleration) in three orthogonal directions. The voltages are converted to digital readings and compared to a baseline which facilitates the exclusion of constant acceleration such as that related to gravity. The digital readings were automatically smoothed and filtered by the Actical software to give aggregated counts of raw perturbations which are equivalent to activity for a given time period or epoch. The epoch chosen for this investigation was 1 minute which is the maximum for this monitor and follows a validated protocol trialled by Hansen et al. [2007]. Henceforth, the number of counts per epoch will be referred to simply as 'counts'.

6.2.3 Data Processing: Standardising Time Monitor Attached

The count levels varied throughout the day so only complete 24 hour periods were included in analyses. Due to reported variation between weekday and weekend day counts (Hansen et al. [2007]), weekdays and weekends were treated separately. For example, if an accelerometer was attached on a Saturday morning and removed the following Friday evening, all data collected on the Saturday and

Friday day would be disregarded because neither day had a complete 24 hour period and weekday and weekend hours were deliberately not amalgamated.

An exception to this rule was applied for the dogs that had their collars removed for known periods (typically overnight) or that had reported times in the car. These time periods were excised from their records but they were still considered to have contributed a full 24 hours in that day. Activity was considered in hourly intervals and, where daily counts were considered, the movement overnight was observed to be very low so effectively it could be omitted without greatly affecting average daily counts.

6.2.4 Data Processing: Time In Car

Three owners gave the specific times when their dogs were in the car. For example, 16:15 to 17:00 on 2nd May 2015. These times were excised from the records of the dogs. Using the three dogs and eight journeys, comprising over five reported hours in the car, an average count level per minute was determined for car travel. A histogram of the counts generated by these car journeys (combined) is shown in Figure 6.2 with the median counts per minute of 57 and a mean of 197. The dogs whose owners who did not give the specific times in the car could not have their accelerometer readings adjusted because it was not known which section of their record to excise. Whether the dog spent time in the car was considered as part of the modelling processes as a binary Yes/No categorical indicator.

6.2.5 Statistical Analyses

The accelerometer data were initially compared to the following factors: whether the dog spent time traveling in the car, whether the exercise was restricted for any reason, the dog purpose, household type, whether there was another dog in the household, whether the dog ran on the lead with the owner, whether the collar was taken off at night, the day of recording, the age group of the dog and the hour of the day. The dog ages were grouped in three different ways. The first involved splitting in six month intervals (1 year - 18 months, 18 months - 2 years etcetera). The second involved splitting into 1 year - 18 months, 18 months - 2.5 years, 2.5 - 3.5 years etcetera. The third was the simplest, splitting into groups above and below 18 months of age.

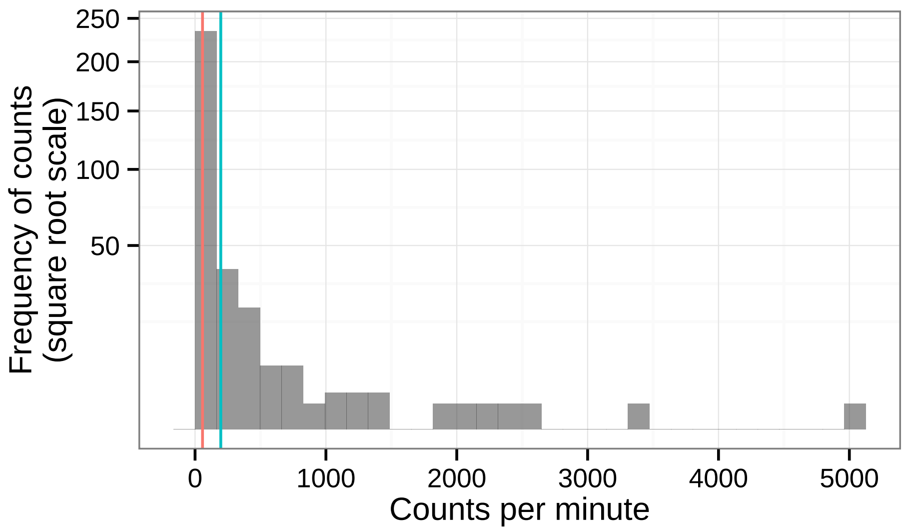


Figure 6.2: Histogram of counts per minute during over five hours worth of car journeys. The mean is given in blue and the median in red.

Table 6.1: Dancey and Reidy's [2004] suggested categorisation for correlation measures

Value of Correlation Coefficient	Strength of Correlation
1	Perfect
0.70-0.90	Strong
0.4-0.6	Moderate
0.1-0.3	Weak
0	Zero

Efforts were then made to develop models of the accelerometer data that included owner reports of exercise. These models were built with a subset of the dogs and then used to predict the accelerometer data for the remaining dogs. Using Dancey and Reidy's categorisation [2004] (Table 6.1), the correlation between model predictions and accelerometer data would give a measure of the validity of the exercise section of the questionnaire and its capacity to predict the accelerometer readings.

6.2.5.1 Dealing with the count data

The count data were considered for every minute as counts and transformed into square root of counts and log of counts+1. They were also grouped into hourly means of the log of counts+1. Michel and Brown [2011] determined thresholds at which they considered activity monitor counts to be associated with sedentary, walking and trotting behaviours in a group of 104 dogs. They characterised these behaviours as indicative of sedentary, light and moderate to vigorous activity and their thresholds lay at under 204, between 204 and 1,751 and over 1,751 counts per minute respectively. These thresholds were applied to the accelerometer data collected by Dogslife such that the activity of each each minute would be considered sedentary, light or moderate to vigorous. For each dog, the percentage of counts that were associated with sedentary, light and moderate to vigorous behaviours were determined for hourly intervals and days.

6.2.5.2 Modelling detail

Logistic regression models using the thresholds were built using the binomial family of distributions in the *lme4* package (Bates et al. [2014]). Three models were built including, for example, a sedentary model whereby the outcome was whether each time point was sedentary or not sedentary. In addition to the logistic models, linear mixed effects models were built with pet ID, previously reported time spent running on the lead with the owner and dog purpose as random effects (*nlme*, Pinheiro et al. [2013]). Again three models were built but this time the outcomes were the percentage of time the counts were above or below the thresholds the previously mentioned. For example, the percentage of epochs when the counts were sedentary (below 204 counts per minute). Multiple models were built and considered. The *MuMIn* package (Bartoń [2014]) was used to find the models with the lowest AIC. Autocorrelation terms were only considered when modelling the hourly mean of log of counts+1. Model assumptions regarding the residuals (normality and homogeneity of variance) were checked by visual inspection of plots of residuals against fitted values.

6.3 Results

Accelerometer readings were collected from 137 dogs between June 2013 and June 2015. One set of readings were incomplete because the dog had chewed the accelerometer and a number of other sets of readings were truncated for reasons that were not reported to Dogslife. There were at least six days of data available for 129 dogs and at least 7 days for 71 dogs. Figure 6.3 shows the seven day plot for a dog chosen randomly from those with seven complete days (ID = 4279). The mean count for this dog was 287 per minute and the median was zero indicating an extremely right-skewed distribution.

The distributions of counts were so right-skewed that illustrating the counts for all dogs was difficult. Figure 6.4 shows the percentage of counts which were zero in each hourly window for all dogs (the minimum number of hours contributing to any hour window was 923). The weekend and weekday plots have slightly different shapes and both clearly show that the dogs were relatively inactive between midnight and six am each day.

Figure 6.5 shows the 50th - 90th percentiles of counts for each hourly period split for week day and weekend days for all 137 dogs. The median count was only above zero for the hour between 9am and 10am at weekends. The 100th percentile is not included in the plot as the values ranged between 4,548 for midnight to 1am at weekends and 32,767 for 9am to 10am on weekdays. Such high numbers dramatically skewed the plot.

6.3.1 Participant Profiles

The dogs comprised 123 household pets, 11 working dogs, 1 assistance dog and 2 described by their owners as having other purposes. They came from households described as more than one adult (62), families (41), single adults (18) and retired (15). Fifty three of the households had another dog and 84 did not. At the time of selection, their most recent reason for exercise restrictions were recommended by breeder or from their own experience (63), owner time (6), dog problem (3), owner ability (1) and unrestricted (64). On average, eight of the dogs ran on the lead for at least 30 minutes each day and a further 35 ran on the lead for less than 30 minutes. The minimum age was 413 days, the maximum was 1,904 days and the distribution is shown in Figure 6.6.

6.3.2 Modelling

Building logistic regression models of, for example, counts associated with sedentary behaviour versus more active behaviour was hampered by the wealth of the data. The models failed to converge when running on both desktop and server machines with error messages that suggested excessive memory requirements. The

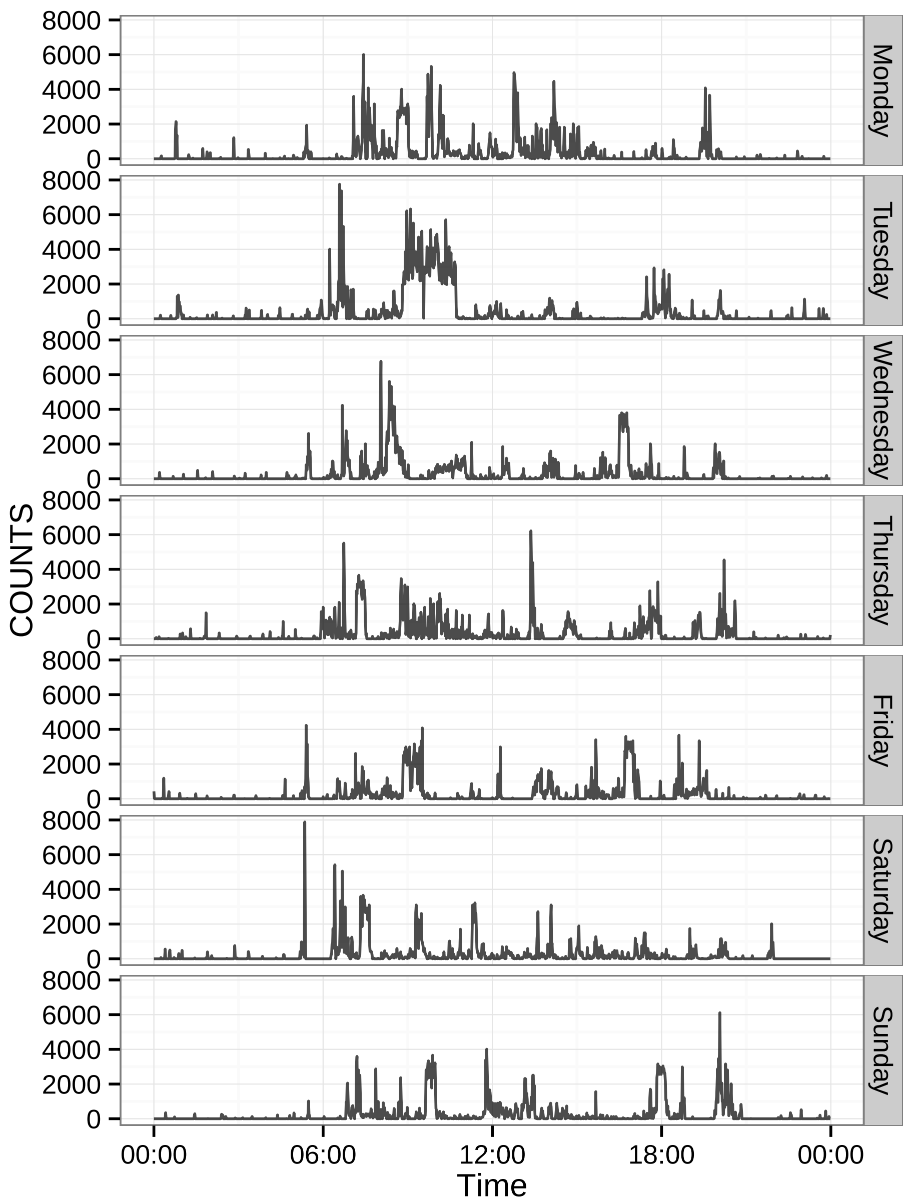


Figure 6.3: Counts for a single, randomly chosen dog, split according to day of the week.

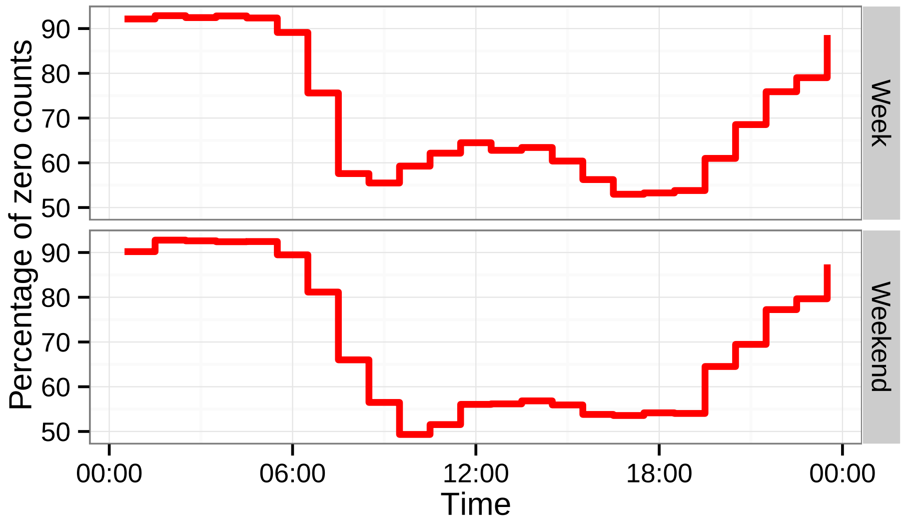


Figure 6.4: Percentage of counts in each hour that were zero

models using count data directly were poorer than those that used the threshold approach due to the extreme right-skew of the data. The threshold approach which circumvented the difficulties of skewed data was therefore applied throughout. The final multivariable models were based on percentage of time spent being sedentary, undertaking light activity and being moderately to vigorously active. As Figure 6.5 demonstrates the counts were dependent on hour of the day and day of the week so both of these elements were included as candidate covariates. Both dog purpose and whether the dog was previously reported to run on the lead were dropped as random effects from the modelling process for reasons which will be discussed below.

6.3.3 Thresholds

On average, the dogs spent 84.8% (95% CI: 83.8-85.8%) of their time being sedentary, 11.3% (95% CI: 10.6-12.0%) engaging in light exercise and 3.9% (95% CI: 3.4-4.4%) engaging in moderate to vigorous exercise. In theory, the non-random sampling should have been accounted for in these averages but the dog purpose was not associated with any of the three threshold values and time spent lead running with the owner prior to sampling was only associated with the time spent on light activity. Dogs that ran with their owner for more than 30 minutes each day spent more time being lightly active than the groups that spent less than

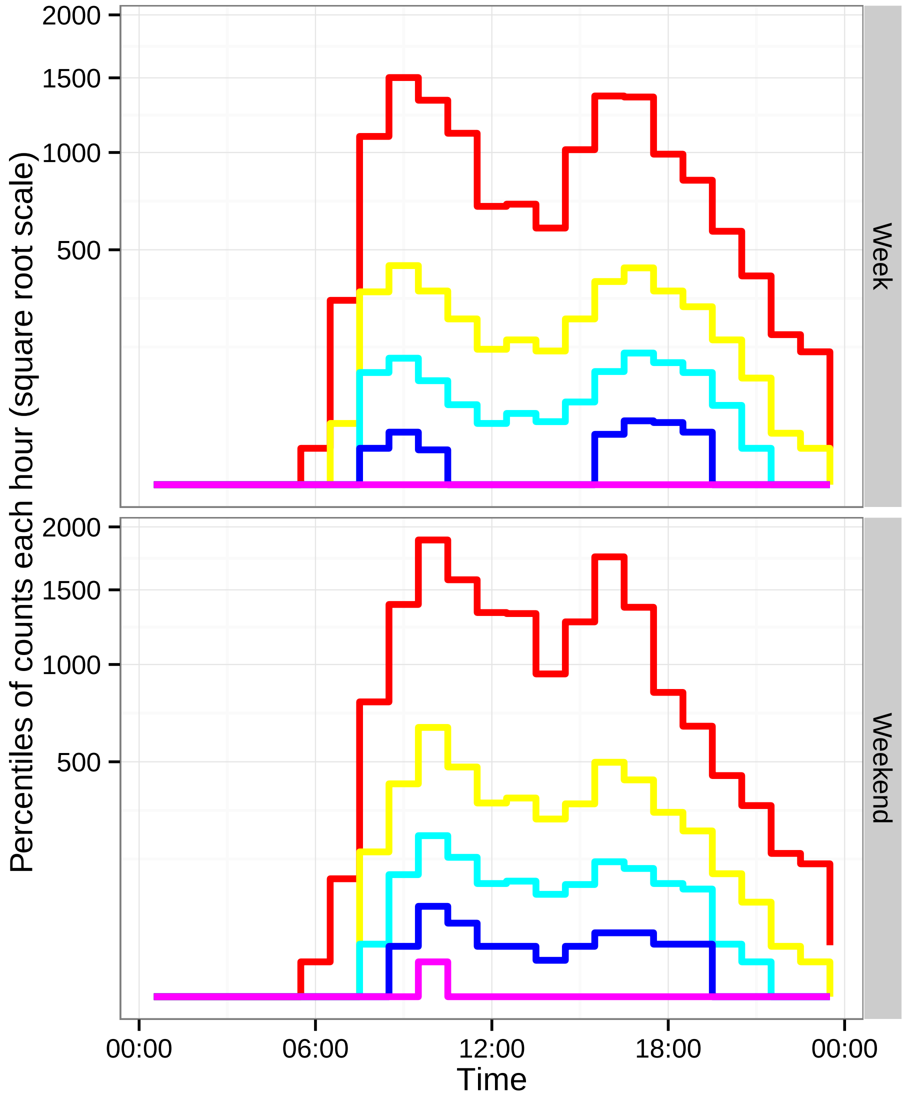


Figure 6.5: Percentile count plot for each hour. The magenta line shows the 50th percentile, blue the 60th, cyan the 70th, yellow the 80th and red the 90th. The 100th (maximum) is excluded because it pushes all other lines too close to the x-axis.

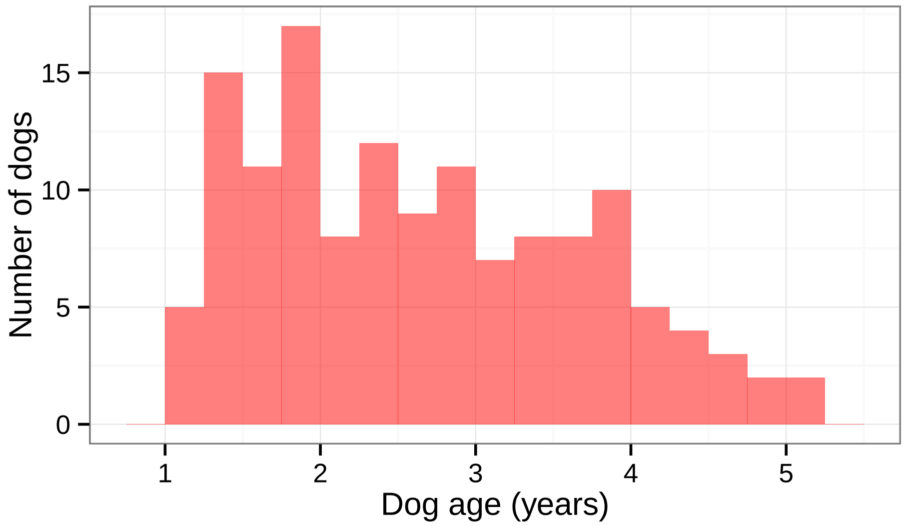


Figure 6.6: Age distribution of the sampled dogs

30 minutes each day (Table 6.3). This was somewhat surprising because, as discussed in Section 6.2.5.1, Michel and Brown [2011] worked out their thresholds for light activity by walking dogs on the lead rather than running.

The averages for groups where differences existed are given in Tables 6.2, 6.3 and 6.4. There were no discernible differences according to whether the dog spent time in the car, having another dog in the household, taking the collar off over night or exercise restrictions. In Section 4.3.11 the amount of reported exercise was shown to increase as the dog aged with dogs over one year reportedly exercising more than those between six and twelve months and more than those under six months. Despite this, the accelerometer readings showed that dogs aged between 12 and 18 months spent a mean of 6.8% less time (median difference = 6.7%) being sedentary than their older counterparts and correspondingly more time undertaking light and moderate to vigorous activity. After 18 months, the increase was a mean of 1.2% per year of time spent being sedentary each day. Dogs from retired households spent less time being sedentary than those from other household types and Saturday was the most active day.

The exercise types that showed any univariable association with accelerometer count thresholds are described in Tables 6.5, 6.6 and 6.7. Lead walking was associated with both sedentary and light activity levels. In broad terms, it appeared that the more time spent walking on the lead, the higher percentage time spent

Table 6.2: Univariable analyses of percentage time spent being sedentary

Variable	Coefficient	95% CI Lower	95% CI Upper	*P* value
Day of the week	83.60	82.50	84.70	
Saturday	0	-	-	-
Sunday	1.05	0.20	1.90	**0.015**
Monday	1.67	0.82	2.52	<**0.001**
Tuesday	1.77	0.93	2.62	<**0.001**
Wednesday	1.17	0.33	2.02	**0.007**
Thursday	1.86	1.01	2.71	<**0.001**
Friday	1.14	0.29	1.99	**0.009**
Age group	78.90	76.40	81.30	
Under 18 months	0	-	-	-
Over 18 months	6.84	4.20	9.48	<**0.001**
Household type	80.90	78.10	83.80	
Retired	0	-	-	-
Family	4.59	1.23	7.95	**0.008**
More than one adult	4.31	1.12	7.50	**0.009**
Single adult	4.15	0.24	8.06	**0.040**

being sedentary but this was belied by the dogs that spent over 2 hours walking on the lead. These animals seemed to spend the least amount of time being sedentary. It should be noted that only dogs that spent 5-15 minutes walking on the lead had a significantly different percentage time being sedentary than dogs who spent over 2 hours walking on the lead. These two categories of time spent walking on the lead also stand out in terms of light activity. Those in the 5-15 minute category spent 2.6% less time on light activity than those who walked for over 2 hours.

Time spent off lead was univariably associated with both sedentary behaviour and moderate to vigourous activity. As might be predicted, the relationships were inverse. The more time spent off lead, the lower percentage time spent being sedentary and higher percentage time spent on moderate to vigorous activity. Dogs that spent over 2 hours off lead each day spent 3% more time doing moderate to vigorous activity compared to those that spent no time off lead and 5.3% less time being sedentary.

Like lead walking, fetching, chasing and retrieving was related to time spent

Table 6.3: Univariable analyses of percentage time spent doing light activity

Variable	Coefficient	95% CI		*P* value
		Lower	Upper	
Day of the week	12.10	11.30	12.90	
Saturday	0	-	-	-
Sunday	-0.67	-1.36	0.02	0.057
Monday	-1.13	-1.82	-0.44	**0.001**
Tuesday	-1.34	-2.03	-0.65	<**0.001**
Wednesday	-0.86	-1.55	-0.17	**0.015**
Thursday	-1.20	-1.89	-0.51	<**0.001**
Friday	-0.51	-1.21	0.18	0.146
Age group	15.50	13.80	17.20	
Under 18 months	0	-	-	-
Over 18 months	-4.84	-6.66	-3.02	<**0.001**
Running with owner*				
(minutes per day)	14.30	13.20	15.50	
None	-3.26	-4.26	-2.27	**0.024**
Less than 30	-3.05	-3.93	-2.17	**0.021**
More than 30	0	-	-	-
Household type	13.30	11.30	15.30	
Retired	0	-	-	-
Family	-2.82	-5.17	-0.47	**0.020**
More than one adult	-2.01	-4.25	0.22	0.080
Single adult	-2.10	-4.84	0.64	0.136

* Whether the dog ran on the lead with the owner was used as a selection criterion for sampling.

Table 6.4: Univariable analyses of percentage time spent doing moderate to vigorous activity

Variable	Coefficient	95% CI		*P* value
		Lower	Upper	
Day of the week	4.32	3.72	4.92	
Saturday	0	-	-	-
Sunday	-0.38	-0.80	0.05	0.086
Monday	-0.54	-0.96	-0.11	**0.015**
Tuesday	-0.43	-0.85	0.00	0.051
Wednesday	-0.31	-0.74	0.11	0.150
Thursday	-0.66	-1.09	-0.23	**0.003**
Friday	-0.62	-1.05	-0.19	**0.005**
Age group	5.62	4.24	7.01	
Under 18 months	0	-	-	-
Over 18 months	-2.00	-3.49	-0.51	**0.009**
Household type	5.73	4.22	7.24	
Retired	0	-	-	-
Family	-1.77	-3.55	0.01	0.054
More than one adult	-2.30	-3.99	-0.60	**0.008**
Single adult	-2.05	-4.12	0.03	0.055

being sedentary and time spent doing light activity. There was a peak of time being sedentary in dogs that spent 1-5 minutes fetching, chasing and retrieving and a similar low point doing light acitivity. If dogs that did no fetching, chasing and retrieving were excluded, there appeared to be a broad trend that the more time spent fetching, chasing and retrieving, the less time being sedentary and the more time spent undertaking light activity.

Time spent doing obedience training appeared to only be associated with moderate to vigorous activity. It was a distinctly non-linear relationship with a peak of moderate to vigorous activity levels in dogs that spent 30-60 minutes daily on obedience training. They nearly doubled their percentage time spent at these high activity levels when compared to dogs that did no obedience training, jumping from 3.3% to 6.1%.

Time spent on other exercise was associated with light activity levels. It appeared that the more time spent on other exercise, the greater percentage time spent undertaking light activity. Dogs that did other exercise for over 2 hours had 2.8% more accelerometer readings at the level of light activity than those that did no other exercise.

Despite efforts to disproportionately sample from the owners who ran with their dog on the lead, no difference was found between the various amounts of time spent lead running.

Table 6.5: Univariable analyses of percentage time spent being sedentary

Variable	Coefficient	95% CI		*P* value
		Lower	Upper	
Lead walking	82.30	79.20	85.30	
None	1.31	-3.17	5.79	0.572
1-5mins	1.36	-2.29	5.00	0.472
5-15mins	4.08	0.70	7.45	**0.025**
15-30mins	2.07	-1.28	5.42	0.235
30-60mins	2.23	-0.94	5.40	0.178
1-2hrs	2.90	-0.08	5.88	0.067
Over 2 hours	0	-	-	-
Off lead	85.90	83.70	88.20	
None	0	-	-	-
5-15mins	-0.26	-3.97	3.46	0.894
15-30mins	-0.24	-2.89	2.42	0.863
30-60mins	-0.45	-2.68	1.79	0.798
1-2hrs	-1.54	-3.85	0.76	0.201
Over 2hrs	-5.34	-8.04	-2.65	$<$**0.001**
Fetching, chasing & retrieving	86.80	84.60	89.00	
None	-1.46	-4.01	1.09	0.27
1-5mins	0	-	-	-
5-15mins	-1.75	-4.06	0.57	0.15
15-30mins	-2.00	-4.46	0.45	0.12
30-60mins	-4.56	-7.44	-1.67	**0.005**
1-2hrs	-3.79	-7.63	0.05	0.064
Over 2hrs	-7.33	-14.50	-0.15	0.056

Table 6.6: Univariable analyses of percentage time spent doing light activity

Variable	Coefficient	95% CI		*P* value
		Lower	Upper	
Lead walking	12.60	10.40	14.90	
None	0.13	-3.17	3.42	0.941
1-5mins	-1.12	-3.82	1.57	0.420
5-15mins	-2.61	-5.10	-0.11	**0.050**
15-30mins	-0.70	-3.18	1.78	0.585
30-60mins	-0.97	-3.33	1.40	0.430
1-2hrs	-1.67	-3.93	0.59	0.157
Over 2hrs	0	-	-	-
Fetching, chasing & retrieving	9.37	7.73	11.00	
None	1.50	-0.47	3.46	0.148
1-5mins	0	-	-	-
5-15mins	1.93	0.17	3.69	**0.041**
15-30mins	1.90	0.03	3.77	0.057
30-60mins	3.85	1.67	6.03	**0.002**
1-2hrs	3.52	0.63	6.42	**0.024**
Over 2hrs	4.50	-1.20	10.20	0.133
Other	10.10	8.35	11.90	
None	0	-	-	-
1-5mins	0.53	-1.81	2.87	0.658
5-15mins	1.05	-0.93	3.03	0.308
15-30mins	1.22	-0.82	3.27	0.253
30-60mins	1.51	-0.61	3.64	0.176
1-2hrs	1.33	-1.14	3.79	0.302
Over 2hrs	2.75	0.42	5.07	**0.029**

Table 6.7: Univariable analyses of percentage time spent doing moderate to vigorous activity

Variable	Coefficient	95% CI		*P* value
		Lower	Upper	
Off lead	3.03	1.98	4.08	
None	0	-	-	-
5-15mins	-0.01	-1.70	1.69	0.993
15-30mins	0.29	-0.90	1.48	0.637
30-60mins	0.42	-0.57	1.42	0.411
1-2hrs	1.37	0.34	2.39	**0.015**
Over 2hrs	2.97	1.76	4.17	<**0.001**
Obedience training	3.30	2.47	4.12	
None	0.87	-0.11	1.84	0.100
1-5mins	0	-	-	-
5-15mins	0.83	-0.19	1.85	0.129
15-30mins	0.19	-1.16	1.54	0.786
30-60mins	2.82	1.02	4.62	**0.007**
1-2hrs	1.33	-0.47	3.13	0.164
Over 2hrs	-0.12	-3.49	3.24	0.944

The final multivariable model for percentage of time spent being sedentary using all dogs is given in Table 6.8. The random effect of pet had an intercept standard deviation of 4.81. When building the model using 65 dogs and predicting the remaining 72 dogs, the correlation between predicted and true accelerometer readings was a moderate 0.43 (Dancey and Reidy [2004]). Surprisingly, despite the inclusion of three of the exercise parameters improving the fit of the model, they did not improve the predicative capacity. The correlation without these factors remained at 0.43.

All three exercise measures that were univariably associated with sedentary behaviour remained in the final, multivariable model.

Table 6.8: Multivariable analysis of percentage time spent being sedentary

Variable	Coefficient	95% CI		*P* value
		Lower	Upper	
(Intercept)	95.70	92.10	99.30	-
Fetching, chasing & retrieving				
Under 5mins	0	-	-	-
5-30mins	-1.13	-2.74	0.49	0.17
Over 30mins	-2.34	-4.46	-0.23	**0.03**
Off lead				
Under 15mins	0	-	-	-
15mins-2hrs	-1.40	-3.56	0.77	0.21
Over 2hrs	-4.86	-7.68	-2.05	**0.001**
Lead walking				
Under 5mins	0	-	-	-
5-15mins	2.59	0.60	4.57	**0.01**
15mins-2hrs	1.06	-0.74	2.87	0.25
Over 2hrs	-0.22	-3.24	2.80	0.89
Household type				
retired	0	-	-	-
family	4.38	1.45	7.31	**0.004**
more than one adult	3.33	0.54	6.13	**0.02**
single adult	3.04	-0.42	6.51	0.09

Continued on next page

Table 6.8: Sedentary multivariable analysis (continued)

Variable	Coefficient	95% CI Lower	95% CI Upper	*P* value
Dog age				
Under 18 months	-5.75	-8.04	-3.46	<**0.001**
Over 18 months	0	-	-	-
Time of day				
Midnight to 1am	0	-	-	-
01:00-	0.67	-1.05	2.38	0.45
02:00-	0.88	-0.84	2.59	0.32
03:00-	0.78	-0.93	2.50	0.37
04:00-	0.63	-1.09	2.35	0.47
05:00-	-1.03	-2.75	0.69	0.24
06:00-	-8.57	-10.30	-6.85	<**0.001**
07:00-	-21.00	-22.70	-19.30	<**0.001**
08:00-	-24.50	-26.20	-22.80	<**0.001**
09:00-	-24.20	-26.00	-22.50	<**0.001**
10:00-	-21.40	-23.10	-19.70	<**0.001**
11:00-	-18.00	-19.70	-16.30	<**0.001**
12:00-	-18.70	-20.40	-17.00	<**0.001**
13:00-	-17.20	-18.90	-15.40	<**0.001**
14:00-	-20.10	-21.80	-18.40	<**0.001**
15:00-	-23.80	-25.50	-22.10	<**0.001**
16:00-	-24.90	-26.60	-23.20	<**0.001**
17:00-	-23.00	-24.70	-21.30	<**0.001**
18:00-	-21.10	-22.80	-19.40	<**0.001**
19:00-	-16.10	-17.80	-14.40	<**0.001**
20:00-	-12.10	-13.80	-10.40	<**0.001**
21:00-	-7.45	-9.17	-5.73	<**0.001**
22:00-	-6.42	-8.13	-4.70	<**0.001**
23:00-	-1.99	-3.71	-0.28	**0.02**

Continued on next page

Table 6.8: Sedentary multivariable analysis (continued)

Variable	Coefficient	95% CI		*P* value
		Lower	Upper	
Day				
Saturday	0	-	-	-
Sunday	1.09	0.17	2.01	**0.02**
Monday	1.59	0.64	2.53	<**0.001**
Tuesday	1.58	0.64	2.52	<**0.001**
Wednesday	1.16	0.22	2.11	**0.01**
Thursday	1.68	0.73	2.63	<**0.001**
Friday	0.98	0.03	1.92	**0.04**

The final multivariable model for percentage time spent doing light activity using all dogs is given in Table 6.9. The random effect of pet ID had an intercept standard deviation of 3.58. When building the model using 65 dogs and predicting the remaining 72 dogs, the correlation between predicted and true accelerometer readings was a weak 0.28 (Dancey and Reidy [2004]). Surprisingly, despite the inclusion of fetching, chasing and retrieving and lead walking improving the fit of the model, their inclusion reduced the predicative capacity. The correlation without these factors was weak to moderate at 0.38.

The relationship between the two exercise categories and percentage time spent doing light activity was not linear. In the case of fetching, chasing and retrieving, when compared to no time spent on the activity, there appeared to be a general finding that some time fetching, chasing and retrieving was associated with more time spent doing light activity according to the accelerometer. However it was only dogs that spent between 30 minutes and 2 hours that were significantly different to the dogs that did no fetching chasing and retrieving. They spent approximately 1.9% more time at light activity levels.

The association between lead walking and light activity levels was also complex. Dogs that walked on the lead for more that 2 hours appeared to have the highest percentage level of time spent on light activity but the difference was only significant when compared to dogs walked on the lead for 5-15 minutes.

Table 6.9: Multivariable analyses of percentage time spent doing light activity

Variable	Coefficient	95% CI Lower	95% CI Upper	*P* value
(Intercept)	2.96	0.65	5.27	
Fetching, chasing & retrieving				
None	0	-	-	-
1-5mins	-1.28	-2.87	0.30	0.11
5-30mins	0.40	-0.86	1.66	0.53
30mins-2hrs	1.92	0.38	3.47	**0.01**
Over 2hrs	4.17	-0.51	8.85	0.08
Lead walking				
None	-0.32	-3.11	2.48	0.82
1-5mins	-0.64	-2.92	1.63	0.58
5-15mins	-2.22	-4.29	-0.15	**0.04**
15-30mins	-0.43	-2.48	1.61	0.68
30-60mins	-1.25	-3.17	0.66	0.20
1-2hrs	-1.51	-3.28	0.26	0.09
Over 2hrs	0	-	-	-
Dog age				
Under 18 months	3.68	2.02	5.35	**<0.001**
Over 18 months	0	-	-	-
Day				
Saturday	0	-	-	-
Sunday	-0.69	-1.32	-0.06	**0.03**
Monday	-1.14	-1.78	-0.51	**<0.001**
Tuesday	-1.28	-1.92	-0.64	**<0.001**
Wednesday	-0.95	-1.59	-0.32	**0.003**
Thursday	-1.11	-1.75	-0.47	**<0.001**
Friday	-0.47	-1.10	0.17	0.15

Continued on next page

Table 6.9: Multivariable analysis of light activity (continued)

Variable	Coefficient	95% CI Lower	95% CI Upper	*P* value
Time of day				
Midnight to 1am	0	-	-	-
01:00-	-0.69	-1.87	0.47	0.24
02:00-	-0.85	-2.02	0.32	0.16
03:00-	-0.71	-1.88	0.46	0.24
04:00-	-0.65	-1.82	0.52	0.28
05:00-	0.94	-0.23	2.11	0.12
06:00-	6.52	5.35	7.69	<**0.001**
07:00-	15.20	14.00	16.40	<**0.001**
08:00-	16.70	15.50	17.90	<**0.001**
09:00-	15.70	14.60	16.90	<**0.001**
10:00-	14.30	13.10	15.50	<**0.001**
11:00-	12.50	11.30	13.60	<**0.001**
12:00-	13.10	12.00	14.30	<**0.001**
13:00-	12.70	11.50	13.90	<**0.001**
14:00-	14.00	12.80	15.20	<**0.001**
15:00-	15.70	14.50	16.90	<**0.001**
16:00-	17.80	16.60	18.90	<**0.001**
17:00-	17.80	16.60	19.00	<**0.001**
18:00-	17.00	15.80	18.10	<**0.001**
19:00-	13.40	12.20	14.60	<**0.001**
20:00-	10.30	9.12	11.50	<**0.001**
21:00-	6.63	5.46	7.80	<**0.001**
22:00-	5.43	4.26	6.60	<**0.001**
23:00-	1.67	0.50	2.84	**0.005**

The final multivariable model for percentage of time spent doing moderate to high activity using all dogs is given in Table 6.10. The random effect of pet ID had a standard deviation of 2.55. When building the model using 65 dogs and predicting the remaining 72 dogs, the correlation between predicted and true accelerometer readings was a weak 0.23 (Dancey and Reidy [2004]). Just one exercise measure, time spent off lead, improved the fit of the model but its inclusion reduced the predicative capacity. The correlation without time spent off lead was 0.26. All of the time categories for time spent off lead were significantly different to the reference category which was over 2 hours and it appeared that the more time spent off lead, the greater the percentage time that the activity monitor would show moderate to high levels of activity. It that respect, it was the clearest of the relationships between one of the Dogslife exercise measures and the activity monitor outputs.

Table 6.10: Multivariable analyses of percentage time spent doing moderate to vigorous activity

Variable	Coefficient	95% CI		*P* value
		Lower	Upper	
(Intercept)	2.55	1.00	4.09	
Off lead				
None	-3.42	-5.21	-1.63	<**0.001**
5-15mins	-3.17	-5.30	-1.03	**0.004**
15-60mins	-2.83	-4.00	-1.66	<**0.001**
1-2hrs	-2.01	-3.16	-0.87	<**0.001**
Over 2hrs	0			
Dog age				
Under 18 months	1.68	0.36	2.99	**0.013**
Over 18 months	-	-	-	-

Continued on next page

Table 6.10: Multivariable analysis of moderate to vigorous activity (continued)

Variable	Coefficient	95% CI		*P* value
		Lower	Upper	
Time of day				
Midnight to 1am	0	-	-	-
01:00-	-0.07	-1.60	1.46	0.93
02:00-	-0.12	-1.64	1.41	0.88
03:00-	-0.16	-1.69	1.36	0.83
04:00-	-0.09	-1.61	1.44	0.91
05:00-	0.01	-1.52	1.53	0.99
06:00-	1.81	0.28	3.33	**0.021**
07:00-	5.42	3.89	6.95	<**0.001**
08:00-	7.74	6.21	9.26	<**0.001**
09:00-	8.51	6.98	10.00	<**0.001**
10:00-	7.15	5.62	8.68	<**0.001**
11:00-	6.03	4.51	7.56	<**0.001**
12:00-	5.37	3.84	6.90	<**0.001**
13:00-	4.82	3.29	6.34	<**0.001**
14:00-	6.71	5.18	8.24	<**0.001**
15:00-	8.73	7.20	10.30	<**0.001**
16:00-	6.87	5.34	8.39	<**0.001**
17:00-	5.18	3.65	6.71	<**0.001**
18:00-	3.67	2.14	5.20	<**0.001**
19:00-	2.51	0.99	4.04	**0.001**
20:00-	1.88	0.35	3.41	**0.016**
21:00-	0.75	-0.78	2.27	0.34
22:00-	0.77	-0.76	2.29	0.33
23:00-	0.25	-1.28	1.78	0.75

6.4 Discussion

There is a wealth of evidence regarding the benefits of regular physical exercise and in 2002, exercise was discussed extensively in a World Health Report regarding the promotion of a healthy life (World Health Organisation [2002]). In order to understand the drivers of wellbeing in the Dogslife cohort, it was important that data were captured that described the dogs' exercise regimes. The Dogslife exercise data were described in Section 4.3.11 and in this chapter, a sample of Dogslife data have been compared with a accelerometer data similar to that collected in multiple other studies (for example, Hansen et al. [2007]; Dow et al. [2009]; Michel and Brown [2011]; Wrigglesworth et al. [2011]; Morrison et al. [2014]). The aim was to determine whether Dogslife exercise data could be correlated with measures used more widely in exercise and activity studies and the degree of success is discussed below.

As described in Section 4.3.11, the exercise section of the Dogslife questionnaire relates to a mean of 157.5 minutes and a median of 128.7 minutes per day. This is just 2.5 of 24 hours per day (10%). During each 24 hours, accelerometers placed on the collar will pick up movement such as that caused by the dog scratching its neck and travelling in a car. These activities were not accounted for in the data collected from owners. Despite this, five of the six Dogslife exercise measures were univariably associated with the percentage of accelerometer counts that met predetermined thresholds and four of the six remained associated in multivariable models that accounted for a variety of other factors.

In addition to the question regarding exercise restrictions, the exercise section of the Dogslife questionnaire offered owners seven categorical time options that related to six exercise types over weekends and weekdays. With accelerometer readings for 137 dogs, perhaps it was optimistic to think that clear associations would be found with such a large set of possible answers. Nevertheless, associations were found between certain activities and percentage time spent being sedentary, doing light activity and undertaking moderate to vigorous activity. For example, a dog that spent over 2 hours off lead each day would have accelerometer readings that showed a relatively high percentage of time doing moderate to high activity levels and relatively low time being sedentary. If the dog also spent over 30 minutes fetching, chasing and retrieving and low, but non-zero, amounts of time walking on the lead, their time being sedentary would drop further. It was a positive finding that answers to the exercise section of the Dogslife questionnaire were clearly associated with accelerometer count thresholds determined in previous studies.

It was surprising to find that time spent running on the lead was correlated with light activity. The original study by Michel and Brown [2011] determined this cut-off by walking with 104 dogs along the same flat path for three minutes each and distinguished this activity from the dog lying down and trotting along the same path. The dogs had a wider range of breed, size and age than the Dogslife cohort so perhaps the threshold levels were not ideal for a cohort of young LRs. The UK-wide nature of the Dogslife cohort make it difficult to determine Dogslife-specific thresholds via similar standardised testing so applying previously published thresholds seemed a sensible strategy. Given that reasonable correlations were found between the thresholds and questionnaire answers, the approach was vindicated. Nevertheless, the correlation between running on the lead and time spent doing light activity remains inexplicable. Perhaps Dogslife owners who run with their dogs for at least 30 minutes run at slower speeds than the dogs that trotted along a flat path for three minutes. Certainly it highlights the difficulties of standardising exercise in uncontrolled settings.

The issue of the age of the cohort was crucial in the finding that dogs under 18 months of age were so much less inactive than dogs over 18 months; a mean difference of 6.8%. The thresholds were originally developed using dogs that ranged from one to twelve years of age and the authors reported that the median percentage of time spent being sedentary increased linearly by 0.9% for each year of age (Michel and Brown [2011]). They did not mention any change in rate between one and two years of age. Given such a large change, in the future, more attention should be paid to the age of dogs contributing to thresholds. Work should also be undertaken to investigate activity in dogs under one year of age in order to complete the picture of how activity changes with age.

The disappointing element of this validation exercise is the predictive capacity of the models that have been built. Including owner reports of exercise according to the Dogslife questionnaire either made no difference or reduced correlations between predicted and true percentages of time at each of the activity levels. This finding may be indicative of model over-fitting but may also be a result of modelling something complex with just 137 dogs. Interestingly, the predictive capacity of the models was best for sedentary behaviour which took up an average of 85% of each day. One might have imagined that the exercise section of the Dogslife questionnaire would best relate to light and moderate to vigorous activity. Instead it appeared that, with the exception of walking on the lead, more time undertaking various activities according to Dogslife correlated with reduced time being inactive according to the accelerometers.

Despite the lack of predictive power with regard to accelerometer counts, the models above do show associations. The Dogslife findings build on work by Morrison et al. [2014] using different accelerometers previously validated by Yam et al. [2011]. They found that dog age and breed were correlated with sedentary and moderate to vigorous activity but just age was correlated with light activity. Neuter status, sex of the dog and body condition score were not correlated with the accelerometer thresholds. The new Dogslife findings mean that the type of household where a dog lives can be added as a correlate to accelerometer readings for sedentary behaviour.

Beyond these relatively stable features, it has now been shown that aspects of the exercise section of the Dogslife questionnaire correlated with accelerometer readings. Previous work has also shown that the exercise section of the Dogslife questionnaire relate to dog weight (Table 4.7) and to a variety of demographic characteristics (Table 4.8). As such, the questionnaire remains a valuable tool for future risk factor analyses about the cohort.

Chapter 7

Limber Tail

7.1 Introduction

In 1997, a series of letters were exchanged in the Veterinary Record regarding dogs with symptoms which the author of the first letter described as “acute onset paralysis of the tail (frozen tail or limber tail)” Hewison [1997]. The first letter detailed cases in two young LRs and one Flat-Coat Retriever and mentioned that onset followed “swimming in, or showering with, cold water”. Subsequent letters by Jeffels [1997], Wilkins [1997] and Steiss [1997b] described similar cases that they had seen and Wilkins was the first to mention that the signs included a “painful tailbase”. The consensus was that the signs typically followed exercising in cold water and that they resolved after a period of as much as 10 days. Wilkins mentioned that the condition was called “Rudder tail” in Norfolk and that the cases they had seen all involved LRs during shooting season.

In 1999, a Norwegian study (Bredal and Thoresen [1999]) for which only the abstract was available in English suggested that limber tail was caused by myositis (inflammation and degeneration of muscle tissue). This was consistent with the findings of a small study of English Pointers (Steiss et al. [1999]) comparing four case dogs with three controls which came from the same kennels. The study involved assessing serum biochemistry, electromyography, imaging (using a variety of methods), thermography and histopathology. All affected dogs had flaccidity of the tail, raised creatinine kinase levels indicative of myopathy and evidence of coccygeal muscle damage.

In a wider report on muscle disorders in working dogs (Steiss [2002]), limber tail was described as a condition characterised by a flaccid tail which either hung from the base or extended horizontally for a short distance before hanging. The dogs would typically recover spontaneously within a few days to two weeks and anecdotal evidence indicated that nonsteroidal anti-inflammatory drugs might speed recovery. The author referred back to their previous works (Steiss and Wright [1995]; Steiss [1996]; Steiss [1997a] and Steiss [1997c]) which apparently included a survey of 113 owners of over 3,000 hunting dogs in the Southeastern USA. The findings from this study have been widely paraphrased online but the original articles are not available as they were published in magazines which cannot now be retrieved. An attempt was made to contact the first author but, as of September 2015, the works remain unavailable. As such, it is difficult to comment on their content and in particular, there is no was of assessing the methodology and any resulting prevalence of the condition in what was apparently a large study.

The first reported prevalence of the condition came in 2008 from a convenience survey of lameness and injury involving over 1,300 working dogs across two hunting seasons in 2005-2007 in Great Britain (Houlton [2008]). Just three of 613 LRs and one of 66 Flat-Coat Retrievers were reported to suffer from 'cold tail' in the study, resulting in an estimated risk of 0.49% in this group of working LRs.

This condition, which will henceforth be referred to as limber tail, appears to be self-limiting, with a low prevalence. Nevertheless, the condition was noted by the Dogslife team as an 'illness' that was being reported to the project with moderate frequency, and it was an issue that had particular relevance to LR. The aim of this chapter was to start using the wealth of illness and lifestyle data collected to focus on a specific illness. Given the large number of dogs involved in Dogslife, even low prevalence illnesses would start to be seen in reasonable numbers. With sufficient reports of limber tail, it should be possible to characterise the condition in the cohort and undertake risk factor analyses. In addition, if sufficient genetic samples were available, it might also be possible assess whether genetic predispositions underlay reports.

7.2 Methods

In Summer 2014, Dogslife data regarding dogs that had a VeNom coded diagnosis of limber tail prior to 31st December 2013 were examined. They did not appear unusual in terms of their demographic characteristics and were split evenly between males and females. A preliminary genetic analysis was undertaken comparing these cases with a group of controls. The controls were chosen from the same group of dogs (discussed throughout this thesis) that had taken part in Dogslife prior to 2014 and they were chosen according to their time at risk in the project. They had to have been part of the project for at least as much time as the cases when their signs of limber tail were reported.

An initial GWAS was undertaken based on 16 cases and controls and multiple promising areas of the genome were identified. There was an awareness that, due to under-reporting, some of the controls might be affected by limber tail but the owner might not have reported their true case status. There was also an awareness that limber tail is typically considered to have environmental causes that would not be captured by the normal Dogslife questionnaire, for example, swimming. As such, more detailed analyses would require extra information. Ideally, the analyses would also be based on a larger sample because 16 cases and controls provides limited power to find associations, both in terms of lifestyle factors but particularly if it was not a monogenetic disorder with complete penetrance.

In June 2014, the Dogslife newsletter included an article about histiocytoma and limber tail in the hopes of generating interest, reports and samples from participating owners (*http://www.dogslife.ac.uk/newsletter/view/48*). The project administrator was contacted by owners of pedigree LRs that were not part of the Dogslife project but who had suffered from limber tail. In order to increase the numbers of samples available for genetic analyses, these owners were asked for genomic DNA in the form of saliva samples. The dogs were typically too old to join the project but belonged to participating members or their friends or family. In addition, colleagues of the project team supplied samples from their dogs.

In May 2015, controls were chosen for all of the cases that had been reported to Dogslife with a VeNom coded diagnosis of limber tail. The aim was to contact the owners of the cases and controls and ask them to complete a questionnaire about signs of limber tail and exposure to swimming. The controls were again chosen on the basis that they had spent as much time as part of Dogslife as the cases (at the time of reporting limber tail signs). In this work, live data were used in order to capture all cases reported to Dogslife in nearly five years of data collection. The potential controls were not chosen at random but instead were preferentially chosen from those dogs that had already contributed DNA samples as described in Section 2.2.4. DNA samples were not available for some reported cases so, if the project administrator was able to contact the owners of these dogs, in addition to asking them to complete a questionnaire, they were also asked if they would provide a saliva sample.

When an owner agreed to take part, they either completed a questionnaire over the phone or were emailed or sent a questionnaire in the post. The questionnaires are available in Appendix 8. The cases were sent one version (Figure A17) and the potential controls were sent another (Figure A18). The aim was to better describe the signs of limber tail and ensure that dogs selected as possible controls were not previously unreported cases.

7.2.1 Statistical Analyses

The numbers of dogs reported to Dogslife with signs of limber tail were characterised. Risks were reported with 95% binomial CIs and rates with 95% poisson CIs, both using *base R* (R Core Team [2013]). The demographic characteristics of these cases and the timing of the limber tail reports were described.

The success of recruiting dogs to the case-control study were described. Dogs initially chosen as potential controls had their case-control status assessed according to the questionnaire answers and, where necessary, controls were re-assigned as cases. The number of different signs associated with limber tail were reported. Exposures reported to occur prior to the onset of signs of limber tail, the number of incidents per dog, the duration of signs and the owner's perception of the dog's pain and quality of life were detailed.

The sex, age, coat colour, height, weight, Dogslife exercise reports, household type, location, owner smoking status and dog purpose were compared between cases and controls. Fisher's exact tests were performed using the *exact2x2* package for 2x2 tables (Fay [2010]) and the *base* package for larger tables in *R*. Weight and exercise were cleaned as described in Chapter 4 but the heights were cleaned manually rather than applying the probabilistic approach. Time spent exercising was treated as a continuous variable created by taking the mid-point of the time categories, as described in Section 4.2.9. Height, weight and exercise have been shown to depend on age and a variety of other factors (Sections 4.3.9, 4.3.10 and 4.3.11). As such, they were initially examined graphically and at specific ages before linear mixed models were built in the *nlme* package (Pinheiro et al. [2013]). These models provisionally included dog age (as a linear and categorical predictor), dog sex, coat colour (in the weight model), dog purpose and case-control status.

In order to assess the degree of relatedness in the cases and controls, permutation testing was undertaken whereby 10,000 samples of dogs the same size as the number of cases and controls were chosen at random from all Dogslife dogs and the number of different sires and dams contributing to these samples was determined. These values were compared with the number of sires and dams that contributed to the cases and controls.

Swimming behaviours were described and answers given for cases and controls were compared. Again Fisher's exact test were used.

7.2.2 Genetic Analyses

DNA was extracted from the saliva samples by Ailsa Carlisle, Roslin Institute. The samples were sent to Edinburgh Genomics and genotyped using the Illumina CanineHD Whole-Genome Genotyping BeadChip. The chip surveys over 170,000 sites on the dog genome, associated with SNPs. The data was filtered to exclude SNPs with minor allele frequency <0.05, genotype call rate <0.95 or those that were not in Hardy-Weinberg equilibrium (significance set at 0.0001). Individuals with DNA samples that failed to type for at least 90% of SNPs were also excluded.

Beyond the individuals mentioned in Section 7.2 above, genotype data were also available for a number of LRs that were not part of the Dogslife project. These dogs had been genotyped as part of other work by the Summers and Schoenebeck groups at the Roslin Institute. There were also a number of Dogslife dogs for whom saliva samples were available but whose owners had not been contacted regarding the limber tail questionnaire. Despite the potential for these dogs to have had limber tail, they were treated as extra controls in a GWAS. The filtering and GWAS were initially performed using Plink 1.07 (Purcell et al. [2007]). A relationship matrix using all of the cases and controls was built using GEMMA (Zhou and Stephens [2012]) and then, also in GEMMA, a GWAS was undertaken as a linear mixed model. The inclusion of the relationship matrix adjusted for underlying correlations in the population structure.

The results were presented as a Manhattan plot to compare with Bonferoni-adjusted significance levels at both genome and chromosome levels. Candidate regions were identified and genes in those regions were examined using the Ensembl browser (*http://www.ensembl.org/index.html*)(release version 81, Cunningham et al. [2014]) and GeneCards (*http://www.genecards.org*). Pubmed was searched using the gene code and the words 'muscle' or 'myopathy' and genes found to be associated with inflammation, mitochondrial function or energy synthesis were reported (*http://www.ncbi.nlm.nih.gov/pubmed*). Typically the genes had been investigated in humans or mice so the information could only be considered as guides to what might be happening in dogs.

7.3 Results

Dogslife received reports of 22 dogs with limber tail in the first three and a half years of the project. In terms of risk, that is 22 of 3,249 dogs with associated data (0.68%; 95% CI: 0.42 - 1.0%) and 1.05% (95% CI: 0.66 - 1.6%) of those with an illness report. If the incidents were considered to be instantaneous, it is equivalent to 7.7 (95% CI: 4.9 - 11.5) incidents per 1000 dog years (considering the total time at risk). Of these, four dogs had two episodes but the remaining 18 had just one. Twelve of the dogs were female, ten were male. Fourteen of the dogs were neutered, and of these, nine were neutered prior to their first instance of limber tail (4 male, 5 female). The dogs were predominantly household pets (20) but there was also one working dog and one multi-purpose dog.

By June 2015, the reported incidents of limber tail had increased to 50 associated with 39 dogs. Thirty-one dogs had one reported incident, six had two incidents, one had three and one had four. There were 22 females and 17 male dogs of which 18 and nine respectively had been neutered although only nine and seven were neutered before their first incident of limber tail. Their coat colours were black (25), yellow (6), chocolate (6), fox red (1) and other (1). They were predominantly household pets (34) but there were also four working dogs and one multi-purpose dog. They came from households that were described as more than one adult (19), families (10), retired (8) and comprising single adults (2).

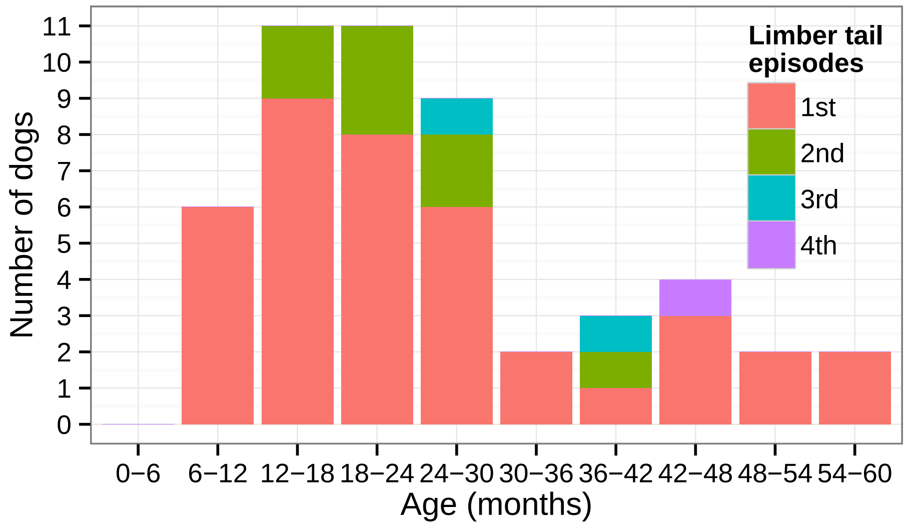


Figure 7.1: Ages of the dogs when they were reported to have developed limber tail

When the dogs had their first reported incident of limber tail, they had a mean age of 25.5 months, median 19.7 months and range 9.6 - 60.0 months. The distribution of ages for all reported incidents is shown in Figure 7.1 and the timing of each is shown in Figure 7.2.

7.3.1 Case-control study

Two hundred and twenty eight dogs were chosen to potentially be part of the case-control study comprising 39 cases and 189 provisional controls. Between June and August 2015, the project administrator attempted to contact 169 owners (33 cases and 136 controls) to ask them to complete the tail health questionnaires. Owners of five of the case dogs had previously been contacted to try to collect DNA but had not replied and the sixth owner had recently mentioned to the project administrator that she was temporarily unable to participate.

The success of the various contact methods are shown in Figure 7.3. In total, 121 of 169 questionnaires were completed (71.6%, 95% CI: 64.1 - 78.2%). As with the collection of veterinary records (Section 3.2.2), those who gave Dogslife permission to contact them by telephone were more likely to get positively involved with 75 (42 + 16 + 15 + 2) of 93 questionnaires completed. The success rate was 80.6% (95% CI: 71.1 - 88.1%) compared to 60.5% (95% CI: 48.6 - 71.6%) for

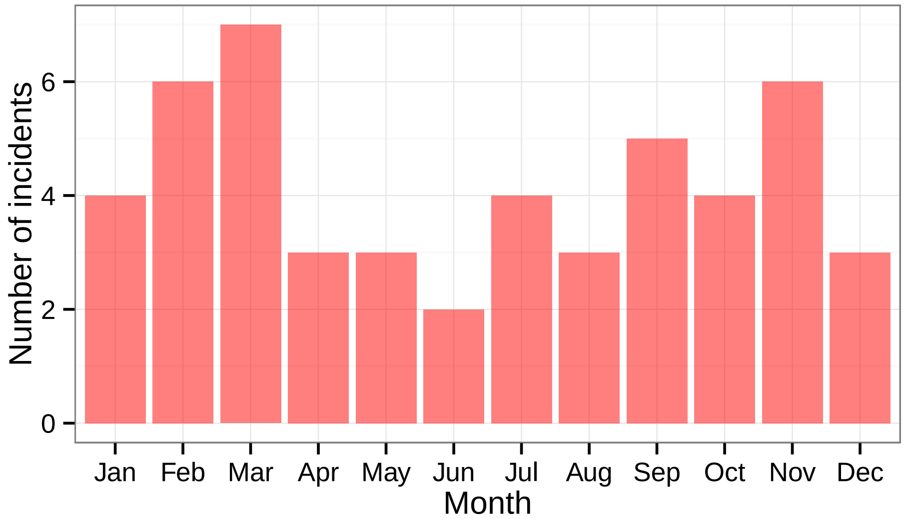


Figure 7.2: Time of year of reported limber tail incidents

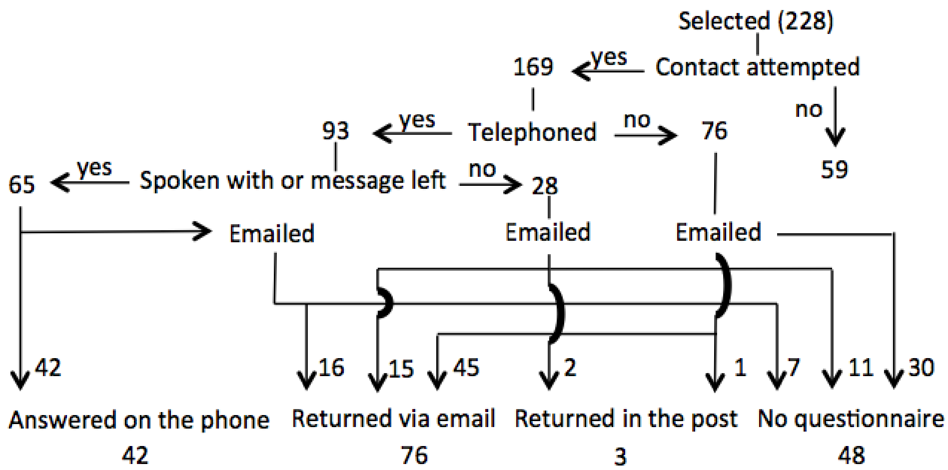


Figure 7.3: Success and failure of obtaining limber tail questionnaire answers

those that did not give permission for telephone contact (irrespective of whether the project administrator were able to reach the owner by phone). Many owners left elements of the questionnaire unanswered so the project administrator recontacted these owners to collect the missing information.

7.3.2 Case Definition

Owners of 29 dogs previously reported to have had limber tail returned their questionnaires and an additional 12 dogs originally chosen as potential controls were reported to have one or more of the signs associated with limber tail. It is fascinating that of 92 potential controls, 12 were unreported cases because this suggests a risk of 13% amongst the cohort, which is far higher than all previously reported estimates. Interestingly, despite the wealth of data, it was difficult to create a case definition based on reported signs. We had previously assumed that the cases would all have limp tails but the owners of three of the dogs (one already reported to Dogslife as a case and two initially selected as controls) did not report limp tails (neither the end nor the entire length). All three had stiff tail bases which might be indicative of other problems not necessarily associated with limber tail. The owner of a dog whose tail was stiff at the base with a limp length said that they believed the dog had a problem with their anal glands rather than limber tail (ID = 171). In the absence of veterinary examination, all 41 dogs that had any of the signs were considered to be cases but analyses were repeated with and without these four dogs that had somewhat ambiguous case-control status.

The details of the signs reported are shown in Figure 7.4 with stiff base of the tail the most frequently reported sign. There was inconsistency from owners who answered ‘no’ to the pain question but then gave a non-zero answer to the pain score. The project administrator explained that the phrasing of the pain question was the issue. The question was “What does you dog’s tail look like when the episodes occur? It appears painful for no reason (Yes/No)”. Apparently for owners that believed their dog had limber tail and thought the tail was painful, this was sometimes answered as ‘no’ because they *knew* why it was painful. Just two owners reported a pain score of zero (no pain) and another three owners gave a pain score of one. By contrast, there were six dogs that had a pain score of ten, two with nine and six with a pain score of eight. The mean pain score was 5.8

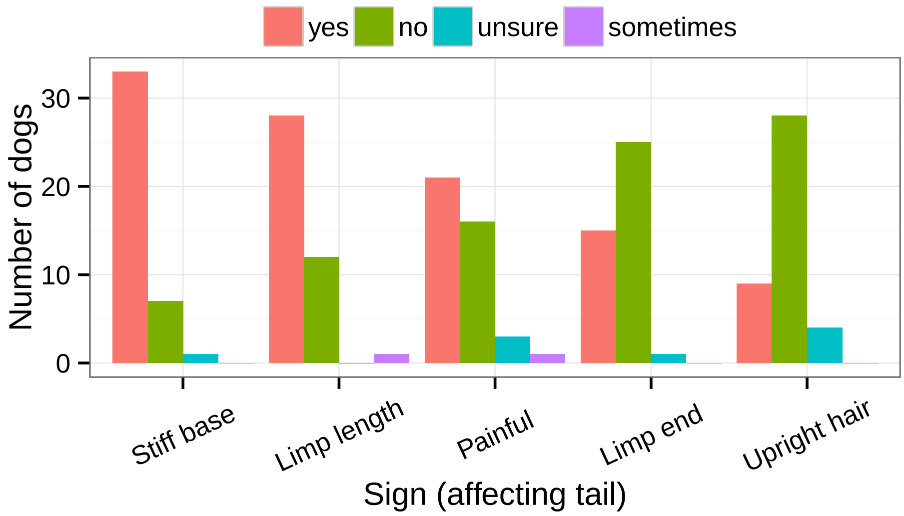


Figure 7.4: Answers regarding signs of reported limber tail

(95% CI: 4.8 - 6.7) and one owner mentioned their dog howling with pain so it was clearly a very painful condition in some cases.

The majority (23) of the dogs had just one episode but nine dogs had two episodes, six had three, two had four and one dog was reported to have 30 episodes. The owner of this last dog had not reported any of them via the online Dogslife questionnaire.

In terms of duration, there was a lot of variation, both between dogs and between incidents for individual dogs. In all four cases where the owner gave different durations for different episodes, the later episodes were shorter. Overall, the shortest episode lasted just a couple of hours and the longest was reported to last about 10 days. If just the first incident for each dog was considered and ‘a few days’ was considered to be three days, ‘a week or more’ was considered to be nine days and a few hours was considered to be 0.25 days then the mean duration was 3.7 days (95% CI: 3.0 - 4.4 days).

Despite owners mentioning relatively high levels of pain, the impact on quality of life was lower with a mean of 3.9 (95% CI: 2.9 - 4.9). The relationship between pain, quality of life, duration of signs and reported number of episodes is shown in Figure 7.5. Duration of signs and reported number of episodes do not appear to explain why some owners who reported high pain scores gave lower scores for impact on quality of life.

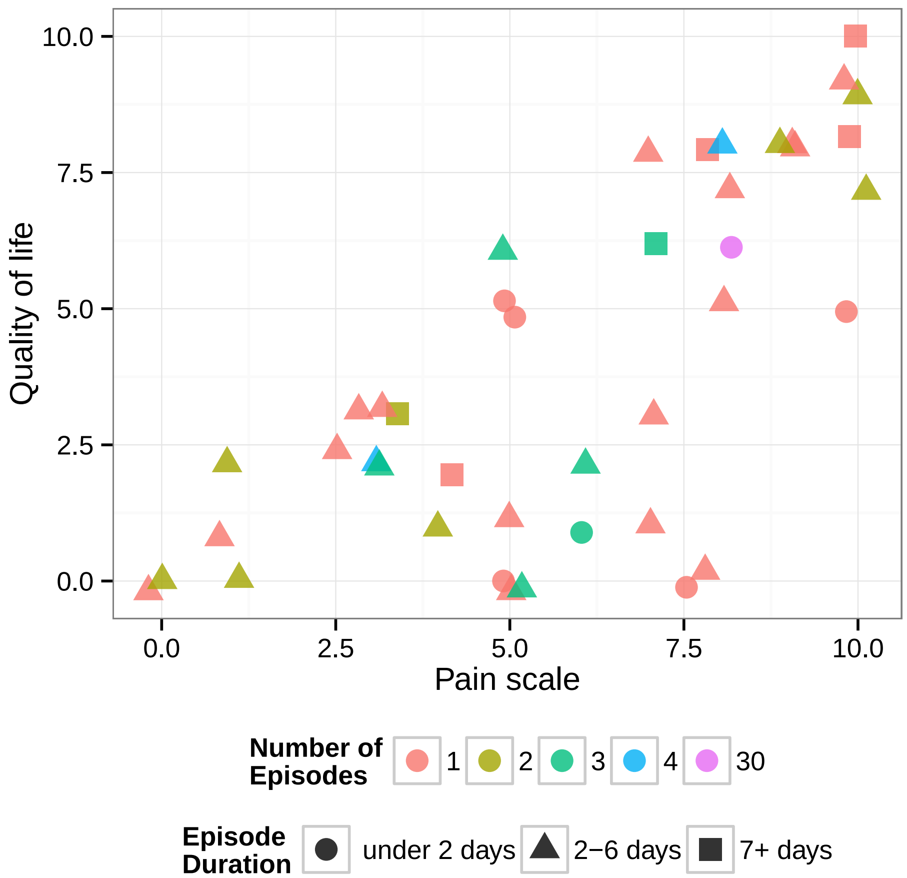


Figure 7.5: Relationship between pain, duration of signs, number of episodes and impact on the dog's quality of life

Owners were asked whether the episodes followed swimming, cold weather, wet weather, vigorous exercise or time confined in a crate. The most frequent precursor was swimming (29), followed by cold weather (20), vigorous exercise (18 and another 3 unsure), wet weather (11 and another 2 unsure), and finally confinement (5 crate, 1 car, 1 unsure). There was no obvious overlap between swimming and cold weather or any other groups of activities. None of the activities was a necessary precursor to limber tail and three of the dogs had done none of the five activities in advance of the onset of signs. To paraphrase these owners, they thought that the signs in their dogs were caused by 'no apparent reason', 'over excited wagging' and 'suspect banged tail when wagging vigorously or maybe reaction to vaccination the previous day'. One of these three dogs was also one of the dogs that did not have any limpness of their tail but the other two both had entirely limp tails. More broadly, 16 of 41 owners wrote something when asked whether they would like to say anything about limber tail. Eight mentioned that they felt it was associated with swimming or water-based play, seven mentioned over-exertion or over-excitement, four mentioned that the dog may have banged their tail and three mentioned vigorous tail wagging.

7.3.3 Case control comparisons

Data were extracted from the Dogslife database relating to the cases and controls assigned via questionnaire as discussed above. The cases comprised 21 females and 20 males compared to 37 female and 43 male controls. The mean age of the cases on 19th August 2015 was 4.41 years (95% CI: 4.36 - 4.46 years) which was lower than the average age of the controls at 4.62 years (95% CI: 4.60 - 4.64 years). This would be expected given that the potential controls were initially chosen on the basis that they had to have been part of Dogslife for at least as long as the cases.

The sires and dams of the cases were considered to see how many parents contributed to the group of 41 dogs. Each case had a different dam but there were just 36 sires; three sires that contributed two cases each and one sire had contributed three of the case dogs. Samples of 41 dogs were chosen at random from the entire Dogslife database 10,000 times, only in 48 instances did a sire contribute three or more of those 41 individuals (0.0048). By contrast, the 80 controls had 75 sires in total with five sires contributing two controls each. One dam contributed two controls to the group of eighty with the remainder coming from 78 dams who each contributed one of their offspring. If 80 dogs were repeatedly drawn at random from the Dogslife cohort then 2,520 of 10,000 samples had dams that contributed two or more offspring. It appears that the cases were more related to each other than might be expected from a random selection but the controls were not.

The coat colours for the cases and controls respectively were black (24 and 42), yellow (7 and 20), chocolate (6 and 17), fox red (2 and 1) and other (2 and 0). They came from households that were described as comprising more than one adult (19 and 35), families (13 and 27), retired (5 and 11), single adults (2 and 7) and unreported (2 and 0) for cases and controls respectively. Within these households, four case dogs and 11 controls came from smoking households compared to 35 cases and 69 controls from non-smoking households (data were missing about smoking for two case dogs). The dogs came from England (31 and 68), Scotland (9 and 10), Wales (1 and 0) and NI (0 and 2) for cases and controls respectively. There were no differences between the cases and controls in terms of these demographic factors.

By contrast, there did appear to be a disproportionately high numbers of working dogs amongst the cases. The cases and controls were described respectively as household pets (35 and 77), working dogs (5 and 2), gundog and pet (1 and 0) and co-counsellor (0 and 1). The odds ratio of a case being a working dog was 5.3 (95% CI: 1.1 - 39.5; Fisher's exact test, $P = 0.04$) compared to not being a working dog. If the four dogs with less clear phenotypes were excluded as cases, the odds ratio increased to 6.0 (95% CI: 1.2 - 44.4; Fisher's exact test, $P = 0.03$) because all of those 'cases' were household pets. If the dog described as a gundog and pet was treated as a working dog and the four unclear cases were excluded then the odds ratio was 7.4 (95% CI: 1.2 - 53.2; Fisher's exact test, $P = 0.01$).

There were 977 cleaned heights and 2,169 cleaned weights available for comparisons. Each dog contributed at least one height and three weights and there were no gross differences between the heights and weights of the cases and controls. After the dogs reached one year of age, the female cases had a mean height of 54.1cm (95% CI: 53.9 - 54.3cm) and the controls had a mean height of 54.1cm (95% CI: 54.0 - 54.2cm). The male cases had a mean height of 59.0cm (95% CI: 58.8 - 59.2cm) and the controls were slightly shorter having a mean height

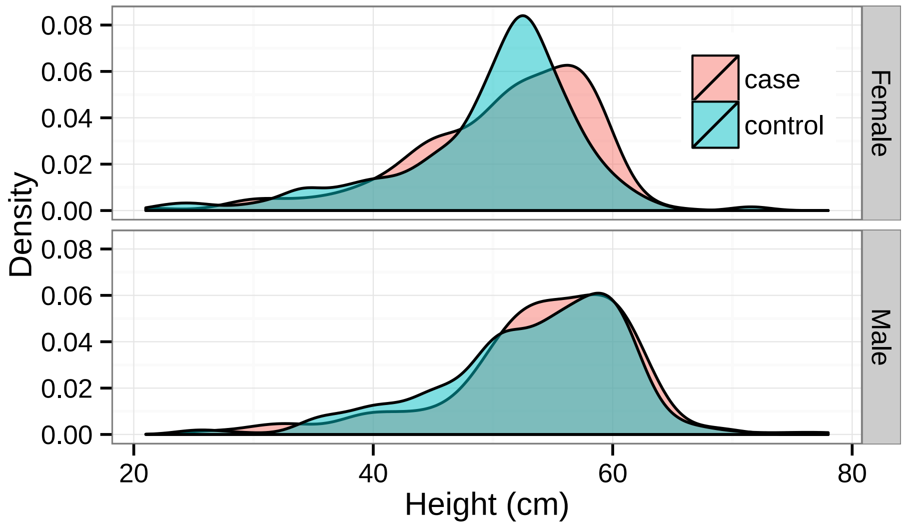


Figure 7.6: Distribution of heights for cases and controls (all ages)

of 57.7cm (95% CI: 57.6 - 57.9cm). A plot of the distributions of 977 heights for all ages is shown in Figure 7.6. Linear mixed models with height as the response variable and including the sex, case-control status and age of the dogs (as both linear and categorical factors) showed no difference between case and control status (details not reported).

The mean weights after the dogs reached one year of age were 26.1kg (95% CI: 26.1 - 26.1kg) for female cases and 26.5kg (95% CI: 26.5 - 26.6kg) for female controls. There was no difference between the means of the male cases and controls; the cases had a mean weight of 31.4kg (31.4 - 31.5kg) and the controls had a mean weight of 31.5kg (95% CI: 31.5 - 31.5kg). The distribution of weights for all ages is shown in Figure 7.7 and, whilst the distributions do not precisely overlap, results of linear mixed models again indicate that there was no difference in weights between the cases and controls.

Across all ages, the mean total time spent exercising for the cases was 173 minutes (95% CI: 148 - 198 minutes) which was indistinguishable from the time spent by controls 162 minutes (95% CI: 131 - 193 minutes). Results of the modelling process indicated that, for ‘other’ exercise and time spent ‘off lead’, the models might be improved by including an interaction term between case-control status and dog age. The details are not reported because only two of over 50 tested interaction terms indicated that the exercise for cases and controls was statistically

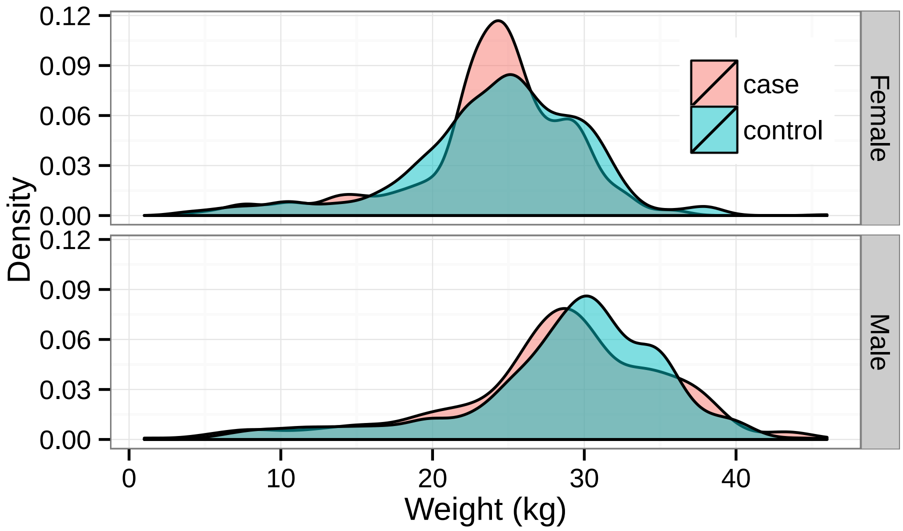


Figure 7.7: Distribution of weights for cases and controls (all ages)

Table 7.1: Associations between limber tail incidence and reported swimming

	All cases		Four cases excluded	
	Case	Control	Case	Control
Swims	38	63	35	63
Does not swim	3	17	2	17
Fisher's exact test	$p = 0.07$		$p = 0.03$	

different at the 5% level. If there was a genuine difference between the cases and controls, it was ambiguous when considering 41 cases and 80 controls, indicating that it was not of overwhelming importance.

7.3.3.1 Swimming analyses

The questionnaire sent to owners of cases and controls included identical questions regarding swimming and, depending on which cases were included in the analysis, it appeared that cases were more likely to swim when compared with controls (Table 7.1). Again, depending on which cases were considered, they were also more likely to only swim in warm water or avoid swimming in cold water when compared with controls (Table 7.2). This ties in with the reports from owners regarding what they do to try to avoid future incidents of limber tail.

All owners who reported signs of limber tail were asked if they did anything

Table 7.2: Associations between limber tail incidence and reported swimming temperature

	All cases		Four cases excluded	
	Case	Control	Case	Control
Warm only	13	10	11	10
All temperatures	25	53	24	53
Fisher's exact test	$p = 0.05$		$p = 0.1$	

with their dog to avoid future incidents. Seventeen owners said they did, sixteen said they did not, four were not sure because their dog had only had one incident and four did not answer the question; this latter non-answer was assumed to be 'no'. Two of the owners who said they did not know how to avoid the incidents nevertheless reported their attempts. The first indicated that when they recognised the symptoms the second time it happened to their dog, they immediately enforced rest and that the symptoms resolved within hours. The second owner mentioned that they kept their dog out of cold water in the hope that it might reduce the risk of future incidents. Beyond these two owners, 17 reported their strategies for avoiding limber tail. This included seven who limited access to water or swimming (specifically in cold water for five of the seven), four who avoided water or swimming altogether in cold weather, three who dried their dog after exposure to water and one owner tried to ensure that the dog swam regularly but for short amounts of time to maintain conditioning and prevent over-exertion. One of the owners who dried their dog also explained that they tried to prevent their dog over-exerting itself. Finally, the owner who said that their dog had a problem with its anal glands rather than limber tail said that they could avoid the symptoms if they got the glands emptied.

A great deal of detailed information were collected about where, how often and when the dogs swam or were exposed to water. The most popular location was rivers (58 of 121) followed by ponds, lakes, lochs or reservoirs (55 of 121) then the sea (51 of 121), then becks, burns, brooks or streams (16 of 121), canals (6 of 121), puddles (4 of 121), hydrotherapy or dog pools (4 of 121) and bogs (1 of 121). The frequency of swimming according to the temperature is shown in Table 7.3 and unsurprisingly the dogs swim less frequently in cooler weather. The answers for cases and controls were statistically indistinguishable from each other at the 5% level but that may be due to low numbers in some categories.

Table 7.3: Number of dogs with reported frequency of swimming according to temperature and/or time of year

	Daily	At least 3 times per week	Up to 3 times per week	More than once per month	Rarely or occasionally
Warm/Summer	20	16	27	8	29
Cold/Winter	13	15	25	14	32

7.3.4 Genetic Analyses

SNP data were available for 142 LRs. They comprised 73 females, 64 males and five for whom the sex was not reported. One hundred and eight of the dogs were part of the Dogslife project and tail health questionnaires had been completed by 75 owners. In terms of cases and controls, there were 42 possible cases (38 Dogslife, 4 non-Dogslife) but this included ID = 171 and the three dogs with no limpness of the tail (see Section 7.3.2). For the GWAS analyses, the case definition was as follows: Reported to Dogslife as having suffered limber tail (either as a Dogslife participant or via the project administrator) *and*, if a questionnaire was completed, there was tail limpness *and* the owner did not have another explanation for the limber tail signs (such as an anal gland problem). This definition was determined through initial examination of the answers to the questionnaire. According to these criteria, there were 38 cases and 104 controls of which 34 cases and 41 controls had completed the tail health questionnaires.

The filtering process reduced the number of SNPs to 111,315 and Figure 7.8 is the Manhattan plot of the GWAS run in GEMMA. These results have been adjusted for underlying relatedness between the dogs.

No SNPs were significant at the Bonferroni-corrected significance level of P <4.49e-07 but there were suggestive hits on chromosomes 6 and 30. Figures 7.9 and 7.10 show the chromosomes in more detail and include Bonferroni-corrected significance lines at the chromosome level.

The genes which have been characterised in those regions were examined and some possible genes of interest are summarised in Table 7.4. Their positions are reported according to the Ensembl CanFam3.1 assembly.

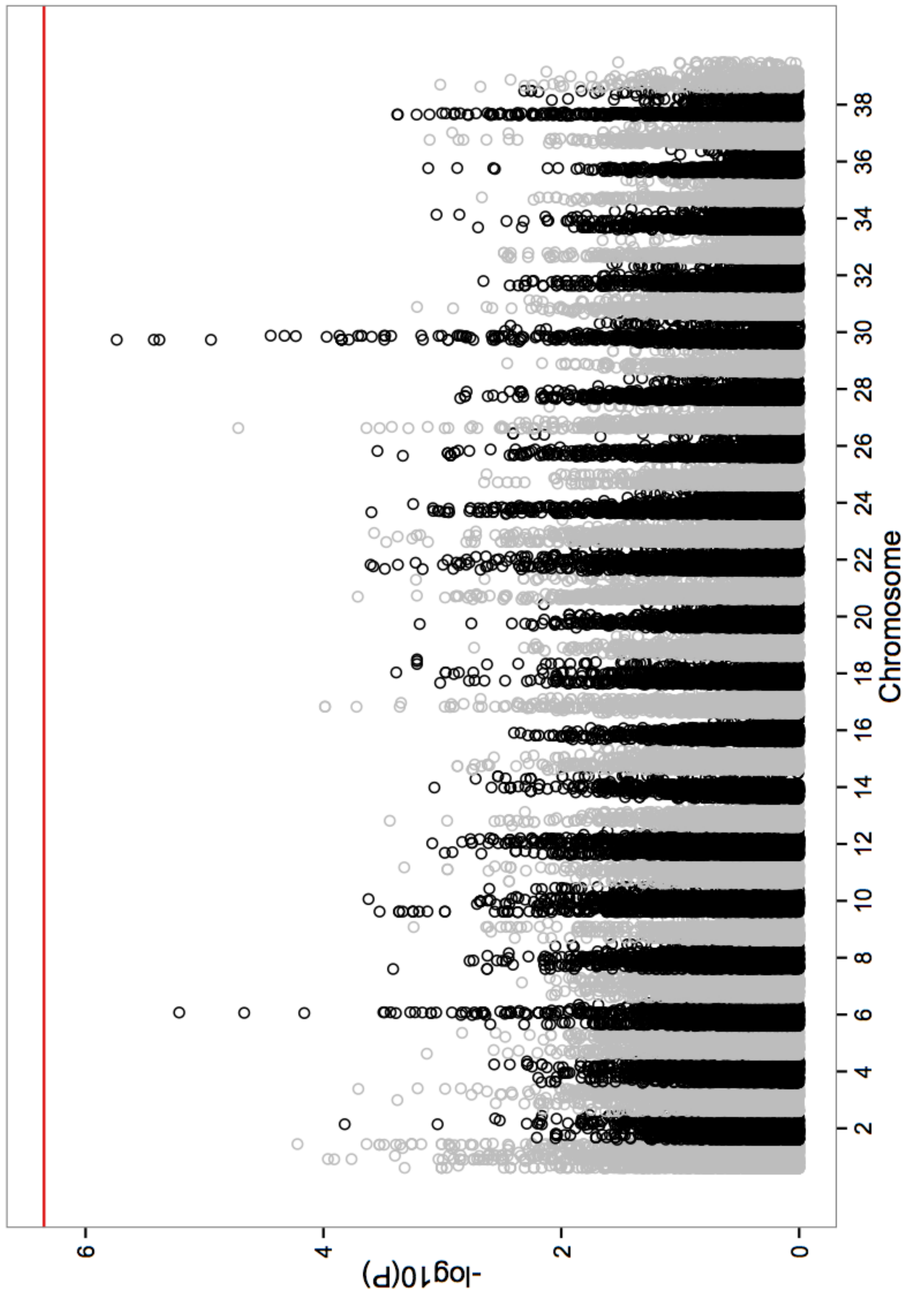


Figure 7.8: Manhattan plot of limber tail GWAS (38 cases, 104 controls). Population structure has been accounted for. Bonferroni-adjusted genome-wide significance line is given in red.

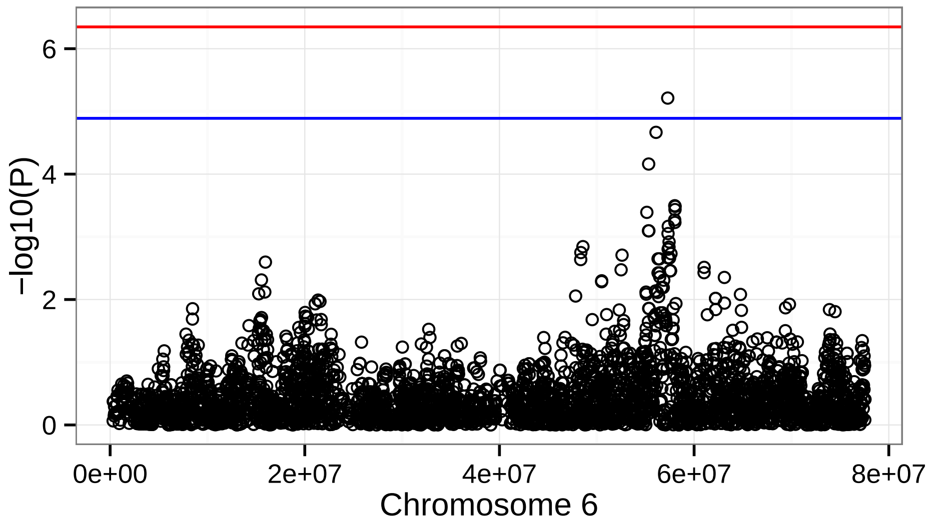


Figure 7.9: Manhattan plot of limber tail GWAS (chromosome 6 only). Bonferroni-adjusted genome-wide significance (red line) and chromosome-wide significance (blue line) thresholds are shown.

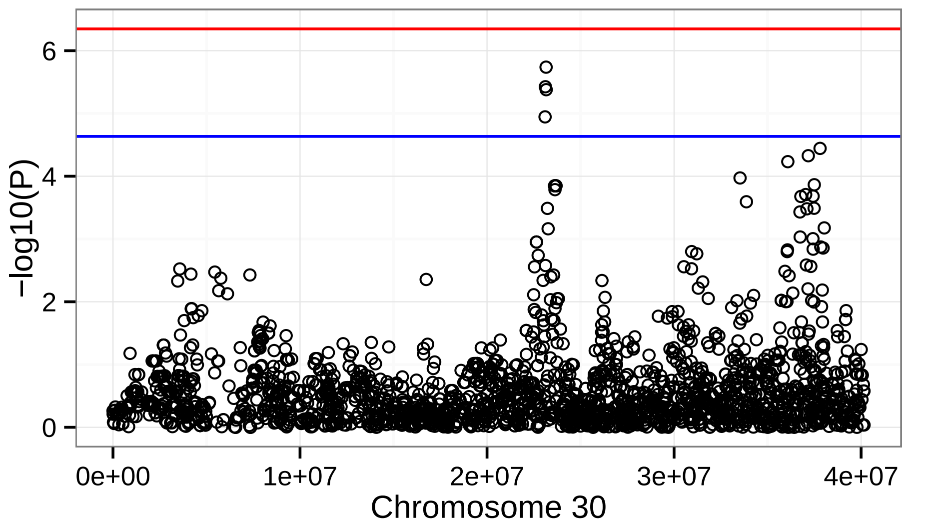


Figure 7.10: Manhattan plot of limber tail GWAS (chromosome 30 only). Bonferroni-adjusted genome-wide significance (red line) and chromosome-wide significance (blue line) thresholds are shown.

Table 7.4: Genes of interest with regard to limber tail

Gene Code	Position	Putative Function
Chromosome 30		
AQP9	23,224,837 forward strand	Encodes membrane protein. Involved in transport of glucose.
ADAM10	23,607,888 reverse strand	Metalloproteinase-disintegrins. Involved in inflammatory response.
TLN2	27,257,989 forward strand	Encodes cytoskeletal protein. Involved in determining skeletal muscle morphology.
ADPGK	36,115,308 reverse strand	Found in mitochondria. Catalyses phosphorylation of glucose to glucose-6-phosphate.
NPTN	36,845,157 reverse strand	Encodes transmembrane immunoglobulin protein involved in cell-cell interactions. Also thought to interact with NDUFV2 which codes for a mitochondrial catalyst.
CYP11A1	37,475,246 reverse strand	Localised on mitochondrial inner membrane and associated with synthesis of lipids.
SEMA7A	37,531,889 reverse strand	Encodes protein that binds to cell surfaces. Promotes production of proinflammatory cytokines.
ARID3B	37,640,641 forward strand	Regulates transcription. Possible role in cell cycle and apoptosis.
CYP1A1	37,793,073 reverse strand	Encodes protein associated with lipid synthesis and iron ion binding.
CYP1A2	37,818,419 forward strand	Encodes protein associated with lipid synthesis. Linked to metabolic pathway and iron ion binding.
Chromosome 6		
GLMN	56,643,620 forward strand	Encodes a phosphorylated protein. Variant associated with development of glomuvenous malformations.

7.4 Discussion

This case-control study is the first example of what might be achieved with the Dogslife cohort for illnesses that are often not presented to veterinarians and which are poorly understood. Swimming has been confirmed as a risk factor for limber tail and working dogs have been shown to be more likely to suffer from the condition. The genetic analyses are not definitive and future work will be discussed below but, at this stage, it is possible to say that limber tail is probably not the result of a mutation of a single gene with full penetrance.

The decision to use a questionnaire to collect extra information and confirm case/control status was important. The data were not collected contemporaneously so they would be subject to all of the issues of a normal case-control study such as recall bias and difficulties with inferring causality but the extra information was invaluable and under-reporting was confirmed. Indeed, an initial risk of 0.68% according to data captured until 31st December 2013 was rapidly superseded by the risk in the provisional controls of 13%. The potential controls were considerably older than the cohort that contributed to the 0.68% risk which included dogs of just a few months of age but 13% was an unexpectedly high value. Whilst it is disappointing that so many of the controls were actually cases that owners had not reported, it reveals a previously unsuspected burden of the condition. Limber tail in LRs is not rare and, given the apparent levels of pain reported, a greater understanding of how to avoid the condition is important.

It is difficult to speculate why owners who are otherwise involved in the project might not report limber tail. Validation indicated that 22% of veterinary presented illnesses went unreported (Section 3.4.4) but it was not possible to assess how many illnesses that were not presented to the vet went unreported. Each illness report takes time so that may be a factor. Beyond time constraints, there was some anecdotal evidence from the project administrator that owners suggested they did not report because they had not taken their dog to the vet. Also, evidence from the validation process indicated that owners did not report if they did not feel that their dog was ill. In this case, owners may have felt that their dog was not ill, it had limber tail (a condition that is easily recognised by owners who have seen it before). These results support refining the Dogslife questionnaire to encourage owners to report anything that affects their dog rather than just those things they consider to be 'illnesses'.

Lifestyle plays a role in limber tail incidence but it was possible to rule out associations with dog height, weight, coat colour, household type, owner smoking status and country location. The relationship with exercise was less clear-cut and, ideally, if more cases and controls can be identified from the cohort then the analyses might be repeated with greater power.

Beyond these lifestyle factors, the potential for a genetic link was illustrated by the finding that one sire contributed three of the case dogs. Initial genetic analyses identified multiple regions of interest and efforts were made to undertake a genetic analysis with increased power. From an epidemiological perspective, undertaking the genetic analyses with non-Dogslife dogs and others of undetermined case/control status was discomfiting. Nevertheless, they were all pedigree LRs and it is standard practice in genetic studies to accept a possible loss of precision in return for increased numbers. It was successful in this instance because all analyses were initially undertaken with just those Dogslife dogs that had confirmed case/control status according to the questionnaires. These analyses were less informative.

The analyses described in Section 7.3.4 have certain limitations. Not all of the cases were defined according to results from the questionnaire. Those without questionnaire details had been reported to Dogslife as limber tail cases but there is no way of knowing whether they had limp tails (the definition used for those with completed questionnaires). As such, some may not have met the ideal case definition. More worryingly, there were multiple issues with the controls. For the non-Dogslife dogs (30 controls) and the Dogslife dogs that did not have an associated questionnaire, no information was available. On the basis that 13% of Dogslife controls were apparent cases, it can be hypothesised that some of these controls were actually cases. There is the additional issue that limber tail has an environmental component and thus dogs could be harbouring the genetic risk without the environmental stimulus to result in the phenotype being expressed. Ideally, a genetic control would be a dog that has spent time working and swimming but still not developed limber tail. Many of the controls did not swim and the majority were not working dogs. As such, they did not undergo the same predisposing conditions as many of the cases. Each of these issues would reduce the chances of finding associations.

Nevertheless, regions of interest on chromosomes 30 and 6 were identified. If the plots had extended above the Bonferroni-corrected significance lines then it would be concluded that limber tail had a clear genetic cause. In this instance, the results should instead be considered suggestive. The peak of each plot is influenced by both the differences in cases and controls but also the underlying allele frequencies. Linkage disequilibrium means that SNPs with a very low P value indicate a region of interest rather than a single gene. The genes under the peaks were examined and those of potential interest are mentioned in Table 7.4.

The investigations into these genes were not exhaustive and relied on the orthologues that could be found in mice and humans. Candidate genes will be briefly discussed below but the next step in genetic investigations of limber tail is sequencing the genomes of four of the cases. The dogs were chosen on the basis that their DNA was of sufficient quality, they had repeated incidents of the condition and that their owners had completed the questionnaire. Sequencing is more expensive that genotyping using the Illumina high density SNP chip (of the order of £2,500 for the four dogs at five-fold coverage compared with £100 per dog for SNP genotyping) but it provides genotyping information for every gene. Unlike the SNP chip which is generalised for use in all dog breeds and therefore does not detect genetic variants in chromosome regions where a specific breed has stretches of homozygosity for the SNPs on the chip, it should capture all the genetic information.

The complete genotype information for the four cases will then be compared with the dog reference genome which is based on a Boxer. Eventually they will be compared with a reference LR (Dog A) which is currently being sequenced at 30 fold coverage and other LRs currently being sequenced at 10 fold coverage as part of work at the Roslin Institute. It will be possible to determine the genotypes for the genes mentioned in Table 7.4 and, more broadly, find genes whereby the four cases have genotypes that differ from Dog A and the other LRs. Analyses can be undertaken to assess whether these differences would likely lead to protein changes and a predisposition to limber tail type signs. For example, GLMN on chromosome 6 can cause glomuvenous malformations in humans which are areas of abnormal tissue that are highly sensitive to changes in temperature. It is not a variation that fits with the Steiss et al. [1999] findings of myopathy, but would fit the phenotype of pain following exposure to cold water. Given that it is unclear how such a variant might present in dogs, it is something that would be interesting to investigate further.

Similarly, there are a number of genes in the region of the peak on chromosome

30 that are associated with changes in mitochondrial function, cell membrane function and the immune response. They all might plausibly be associated with the variety of signs seen by owners of dogs with limber tail. Currently, there is no indication of which gene is a more likely candidate nor of why they might specifically cause myopathy or myositis in the tail. Work assessing the variants in the four fully genotyped dogs should help elucidate these issues.

This short case-control study is far from complete but it has yielded some very interesting findings and there are a clear series of steps ahead to complete the investigation. It is hoped that further work by the Dogslife team will enable the identification of genes associated with limber tail. Armed with this sort of information breeders of working dogs might decide to avoid breeding from dogs carrying the risk variants. The chances of future generations developing this painful and distressing condition might be reduced which would be an wonderful return for the thousands of owners who contribute to Dogslife in the hopes of improving the health of LRs in the UK.

Chapter 8

Discussion

8.1 Introduction

At its inception, the Dogslife project was unique in canine health. Prior to the wide-scale availability of the internet, many questions regarding canine epidemiology were neglected because financially, addressing them was impractical. *Where do dogs live? What do they eat and how much exercise do they get? How big are they?* Crucially, *what illnesses do they develop?* If it was possible to collect appropriate data to answer these questions, how would all of these different inputs combine with the underlying genetics of a pedigree cohort to answer the key question: *How should dogs be bred and raised in order to achieve a healthy canine population?* Dogslife was a pilot project that aimed to try to find out whether it was possible to answer these questions by recruiting dog owners and asking them to repeatedly complete an online questionnaire. As the first PhD student to work on the project, I had the opportunity to fully engage with the data collected during the first three and a half years of Dogslife. It has given me a unique perspective to comment on its strengths and weaknesses and suggest improvements for the next iteration of the project. I will discuss what I have found and what I would like to do next in a difficult financial climate before finishing with ideas of approaches that could be used if money were no object. This thesis is the product of dealing with real world challenges so it is appealing to conclude by taking things into the realm of the science fiction.

8.2 Dogslife In Context

In this thesis, the cohort of owners is well characterised. It is typical of self-selecting studies that the participants have higher than average educational attainment (for example Sullivan et al. [2011]) but this type of information was not collected from owners during Dogslife so this aspect of selection bias could not be assessed. As has been found in other studies where participants are self-selecting (for example, Søgaard et al. [2004]), Dogslife owners are disproportionately female (Section 2.3.3.4) but otherwise seemed representative of eligible owners. It is impossible to assess what impact the differences between those who participate and those who do not register might have on the life course of their dogs but the coat colours, geographic location and sex of the dogs in the cohort were in proportion to those eligible to join. It has been broadly assumed by the Dogslife investigators that the dogs in Dogslife are representative of the wider pedigree LR population in the UK.

Of more concern was retention bias because owners with different demographic characteristics were being disproportionately lost to the project (Table 2.5). Owners who smoked or who were part of family households were most likely to leave the project and owners who had another dog or who lived in retired households were most likely to be retained. The ability to contact owners by email and telephone improved retention but after three and a half years, only a third of owners were up to date. There was an initial loss between registration and answering the questionnaire and then an ongoing loss to the project (Section 2.3.4). Despite the possible consequences of such losses, Dogslife findings are not invalidated but rather must be reported in context. Certainly Dogslife is not alone in losing participants longitudinally.

Across human literature, a wealth of studies address retention strategies in different groups regarding different types of outcomes. There is an awareness that disproportionate loss to follow-up can bias findings from studies and that every effort should be made to minimise such damage. One large-scale study of 25,000 USA students lost 51% of their participants over a 12 year period (Curtin et al. [2002]). As mentioned in the introduction (Section 1.1), the ALSPAC study recruited pregnant women and followed their children as they aged (Golding [1990]). One of the many uses of the cohort involved questioning teachers in the Avon area when the children were nearly eight years of age (Wolke et al. [2009]). Behavioural traits of children who were still participating were compared with children that had been lost to the project. The initial ALSPAC cohort comprised 14,062 live births and after eight years, addresses were available for 10,431 children. The behavioural study was based on 3,946 retained members (38%) and 1,130 lost members (11%) which was just 37% of those believed to be alive and 49% of the members for whom address details were available. Like Dogslife, they found that drop-out was associated with smoking (in this case, maternal smoking) and also that the children who were lost to follow-up were disproportionately likely to suffer from disruptive behavioural traits. If prevalences of behavioural problems were based just on the continuing participants, they would have been significant underestimates.

Beyond human medicine, there are a number of animal studies which Dogslife can be compared with. A study of Boxer dogs in the Netherlands which asked owners of over 1,800 puppies to complete questionnaires every six months found that 32% of their owners failed to complete the final questionnaire when the dog was four years of age (Nielen [2000]). The authors did not report demographic comparisons between owners that left and remained in the study but stated that they did not believe their study was subject to bias.

Results from a study of four large breeds of dogs in Norway have been mentioned throughout this thesis (Indrebø et al. [2007]; Krontveit et al. [2010]; Sævik et al. [2012]). The methodology involved giving owners diaries and asking them to complete questionnaires when their dogs were three, four, six, twelve, eighteen and twenty four months of age. At 24 months, retention was at just over 51%.

In the UK, Bristol University are running a longitudinal study of cats *(http://www.bristol.ac.uk/vetscience/research/projects/cats/)*. They send owners a questionnaire at registration and then when their cat is six, twelve, eighteen and thirty months and at four years of age (Murray et al. [2015]). Their retention was reported to be 87% and 79% at six and twelve months respectively. A number of those lost were reported by owners to have died, been re-homed or gone missing.

It is disappointing to note that Dogslife retention is poorer than for these other studies mentioned and it is likely that multiple reasons underlie the difference. Human studies typically have the advantage of greater funding which facilitates publicity drives and enables the employment of people who can seek to trace participants who lose touch with a study. For example, one ALSPAC publication describes tracing participants who had not contributed in the previous five years by health records, directory enquiries, the electoral register and a variety of other publicly available resources (Bray et al. [2015]). Such an effort would be well beyond the financial capacity of Dogslife and it also relied on linkage with primary and secondary health care records via the UK National Health Service (NHS).

In the field of animal health, the longitudinal studies mentioned were not based entirely online and it is unclear how intermittent technical problems (with the Dogslife website and at the user end) might affect retention. The other studies also asked for data less frequently than Dogslife. It is clearly a trade-off. Validation work in Chapter 3 indicated that frequent questionnaire replies were crucial for monitoring growth in the young dogs and also for limiting recall decay with regard to vaccinations and illnesses. The result of asking owners to return to Dogslife so often appeared to be a higher attrition rate. There has been some discussion amongst the study team regarding whether Dogslife should move to less frequent questionnaires but the potential loss of health information means that, as of August 2015, the study design remains unchanged in this respect. The attrition is well characterised and whilst it is clearly undesirable, we believe that it is worth pursuing frequent questionnaire answers in order to maximise illness reports.

In Chapter 5 these illness data have been explored. Diarrhoea and vomiting reports across the UK were characterised adding to the findings from Norway regarding four large breeds of dog, including LRs. The complications of the Dogslife study population being spread across multiple countries within the UK made geographic comparisons harder but it was still possible to determine an increased hazard of diarrhoea associated with greater human population density (Section 5.3.3). More broadly, it was shown that the peak rates of gastrointestinal signs, musculoskeletal signs and wound or trauma occurred at different ages in the cohort (Figure 5.2).

In the absence of a large increase in funding, many of the approaches employed to improve retention in human studies are beyond Dogslife. However there is a substantial group of owners who register but go no further. It is possible that providing more information to these people in advance of registration might better prepare them for the questionnaire ahead. The project might recruit fewer people but more of them might be retained. Obviously this would entail a different participant profile but, as previously discussed, Dogslife participants are a self-selecting group and are therefore already non-random.

Beyond the short-term attrition, there is a gradual ongoing decline in numbers. Some dogs were re-homed, some owners became unwell or sadly died and some owners simply lost interest. Perhaps when they joined the project, they did not appreciate that the aim was to follow the dogs throughout their lives. As a goal this was ambitious because funding was initially only available for website development and two years salary for the project administrator. There is a new project that started recruiting in 2013 in the USA with the intent of studying 3,000 GRs and the project website clearly states that their aim is to recruit for a 10-year study (Guy et al. [2015]). I will be intrigued to see whether they have a similar level of loss as Dogslife. Potential participants might be put off by such a long-term commitment and this might be reflected in recruitment of a different type of participant who would then be more likely to stay with the study.

The GR study also involves collecting lifestyle data including information regarding physical activity but, to date, their findings remain unpublished. Much like Dogslife, there will be a lag whilst they recruit enough dogs and begin to deal with their data.

To my knowledge, Dogslife is the first study to try to characterise canine lifestyle in the UK. The exercise findings were particularly interesting and again highlighted geographical differences in the cohort. Dogs from Scotland and Wales were reported, on average, to spend more time exercising than those from England (Section 4.3.11). These differences, whilst not enormous, remain in the multivariable model and might reflect cultural differences between the people in the three nations.

In the absence of detailed information about the GR study, comparisons are difficult but they apparently have the funding to collect DNA from all participants. Unfortunately Dogslife funding is more limited. Samples have been collected from a substantial subset of the cohort (Section 2.2.4) but each pair of saliva and faecal samples cost approximately £10 to collect. Ideally the study team would have recovered a DNA sample from each dog registered with the project, to ensure there was an archive of dogs which might subsequently leave the project. Dogslife dogs have already been lost to neoplasia and trauma and the opportunity to genotype these dogs has been missed.

8.2.1 Dogslife and Veterinary Involvement

The methodology of the GR study involves contact with the veterinarian of each dog and they offer owners an incentive of $75 per year to help ensure that they visit their veterinarian for check-ups. Veterinary involvement is something that the Boxer dog study in the Netherlands and the study of four large breeds in Norway also included. It was considered by the Dogslife study team but it was felt that this would create more problems than it solved. If Dogslife were to collect data directly from veterinarians then they would need to recruit owner-veterinarian pairs and this would likely have considerable impacts on recruitment. The project team instead attempted to collect veterinary information by asking owners to take the DHR (Appendix 4) with them each time they visited a vet. It was unfortunate that compliance regarding the DHR was poor but heartening to note that, when the owner reported an illness, the validation process found owner reports and veterinary records were entirely consistent (Section 3.4.4).

If owner-veterinarian pairs could be recruited then there would be more avenues available to maintain contact for retention purposes but it would introduce new practical challenges. Vets would need access to the Dogslife website and we would need to develop a method whereby each report they made would be linked to the correct dog. In the absence of the online facility, completed facsimile or postal versions of the DHR would need to be sent in by veterinarians and someone at Dogslife would have to add them to the relevant dog's profile. It might have been possible to automatically collect electronic records in a fashion similar to that done by VetCompass (VetCompass [2014]) but that is a huge undertaking involving multiple different practice management systems. Effectively it would involve setting up two projects: Dogslife *and* VetCompass.

There is no doubt that veterinary involvement would give greater diagnostic surety and that if automatic record assessment were undertaken then recall decay would be minimised. However, the issue of illnesses that are not presented to vets would remain. Limber tail was investigated in Chapter 7 and it is an illness which is often not presented to veterinarians. In terms of diagnostic clarity, it has signs that are limited in duration and may be confused with spinal damage, a broken tail and issues with the anal glands. For this sort of illness, the availability of veterinary records would often add nothing and if the case definition was based on veterinary diagnosis, the majority of cases would be missed.

Recruiting owner-veterinarian pairs into Dogslife would likely result in a study involving smaller numbers akin to those previously mentioned in Norway, the Netherlands and now the USA. Dogslife applies a new methodology and as this thesis demonstrates, health information reported by owners is sufficient to identify new risk factors for illnesses such a vomiting and diarrhoea and to undertake specific case-control studies such as that regarding limber tail. Dogslife is a pilot project which was set up and run with limited financial resources. Under those constraints, the data collected were remarkable.

8.3 Limitations and suggested improvements

Dogslife owners have contributed invaluable data about their dogs but the data were not easy to deal with. The database where they are stored was designed for use in epidemiological studies but the implementation of the design was not ideal. Entries in separate tables should have been linked by unique IDs but there were multiple duplications and spurious entries. IDs that should have uniquely linked to a single datum in another table might be linked one to many and many to one. Some of the data issues had patterns that were possible to understand and deal with, for example, the modelling developed to deal with the cm/inches issue for heights (Section 4.2.4). Unfortunately, many issues were only identified by accident and remained inexplicable. "There shouldn't be that many" became a mantra. The ultimate solution would have involved viewing each individual entry, then cleaning and re-entering it in a new database. At the time of writing in August 2015, the database comprised 2.7 million rows of data so row by row checking would not be feasible.

With hindsight, attention should have been paid to the database when data began to be collected. Issues with the website were obvious because it was repeatedly tested by the project team and because, after launch, owners would get in touch to report issues. Problems with the database only became apparent when analyses were attempted many months after launch. Each new investigation meant identifying new ways for the data to be inexplicable. By that stage, the person who built the website and database was no longer working for the external contractor and the underlying IT infrastructure was ageing. Fixing one problem seemed to inevitably create another.

Beyond implementation problems, there were other issues with the website which affected the quality of data available for analysis. When designed, it was intended that if an owner mentioned an illness was ongoing, they would be asked about this illness when they returned to the site. Unfortunately this functionality only worked for the first few months of the project. The break-down was not noticed for some time and has never been fixed which means that, unless the illness had resolved before the owner entered the data, an illnesses would have no end date (Section 5.3.1).

In the future, automatic logic checking could be applied to all dates entered by owners relating to illnesses, neutering and vaccinations. A pop-up could appear if they attempted to enter something illogical. This would prevent owners entering dates before the dog was born or in the future. It would also prevent owners creating illnesses that lasted for negative amounts of time. Similarly for heights and weights, sufficient information is now available to suggest a range of possible heights and weights for all ages. It would be more complicated but still possible to apply this to food weights. One imaginative owner reported that their dog was fed 6.02 x 10^{23}g each day (Avogadro's constant) and, whilst this was entertaining to read, flagging the measure as improbable as the owner tried to enter it might have encouraged them to enter the true weight of their dog's food.

For some aspects of the questionnaire, it was unclear whether an owner had answered "no" or simply failed to answer the question. This applied to elements of the registration process, questions about preventative veterinary care and the illness section of the questionnaire. Ideally, if the owner answered "no" this should be recorded to distinguish between those who say "no" and those who have left the website.

Resolving these issues would have helped analyses undertaken by the Dogslife team but there were also errors that impacted on owners. The website needed to work on diverse platforms, for thousands of different owners and there were times when errors occurred. It is likely that many owners did not take the time to contact Dogslife but the project administrator kept a record of all issues reported. In the first three and a half years 173 separate problems were noted and a number of specific issues were identified that had widespread impact. For example, there was an intermittent error that meant owners could not move directly from registration to the questionnaire, presumably resulting in some owners leaving the project and never returning.

A section relating to mating and breeding was added to the questionnaire in June 2014 and suffered from an error that made it unusable until March 2015. During this time, owners became increasingly frustrated because data they entered was not being recorded and they were repeatedly asked to re-enter it.

More broadly there were issues that developed as the size of the cohort increased. The front page of Dogslife included a map which showed where each dog was located across the country. As the number of locations increased, this map took longer to load, to the extent that many owners could not load the front page of the website. If they were able to get past this page and login, there was another issue with the front page of each dog's profile. There was a plot of the dog's height as it aged which included the mean for Dogslife dogs at each age. The script which worked out these averages became slower and slower and many owners would be timed out of their session without ever seeing the front page of the profile. It was frustrating for owners and exasperating for the project team because we did not have the technical expertise in-house to fix these problems.

The many and varied issues with the website and database are not exhaustively listed here. Suffice to say that despite a great deal of expertise and thought being involved in the design and implementation of the data collection process, the data are not ideally suited for epidemiological analysis. To a large extent, this is due to the complexity of the Dogslife questionnaire and the limited funds available for building and maintaining the website and database. Every member of the project team has worked incredibly hard to keep Dogslife running despite the issues. Funds have been secured to update the website and database which will enable to the team to build on all of the work done to date.

Human behaviour has changed since the Dogslife project began. More participants now access the website via a smart phone and, in an ideal world, there would be Dogslife apps for Android and Mac OS. It would offer greater convenience and immediacy; owners could answer the questionnaire whilst standing in their veterinary surgery. At the time of writing in August 2015, the Dogslife website is being rebuilt. The aim is to make the website accessible via multiple platforms (phone, tablet, laptop and desktop). It will not have the simplicity of an app but it must be recognised that the Dogslife website is not simple and the cost of creating apps for multiple platforms is extremely high. Owners are able to create scrapbooks, upload photographs, view the newsletter and featured dog archives *and* answer a complicated questionnaire. Whilst a questionnaire app might be desirable, the ongoing financial commitment involved in keeping it secure would be prohibitive. Much of what we have learned over the the past five years is being implemented in the new version of Dogslife and there is optimism that the new Dogslife website and database will be easier to manage.

8.4 Wider utility of Dogslife

As a unique resource, the Dogslife cohort have already been used for studies not reported here. Collaborations with Cambridge University have included promoting an obesity study called “Go Dogs” to the Dogslife cohort. DNA samples for 386 dogs were sent to Cambridge and the frequency of an allele of interest has been characterised in the LR population. As a consequence, detailed dietary questionnaires will be sent to approximately 60 Dogslife participants. There was a separate request to access the faecal samples but this foundered when the collaborators failed to agree legal terms regarding sample handling.

Internally, there have been two undergraduate student projects involving photographs collected by Dogslife. The first assessed the possibility of detecting dental disease at one year of age and the second was a visual assessment of body condition and obesity. A further project screened 24 faecal samples for 12 viruses and all 372 faecal samples are now being sequenced for 16S bacterial DNA. We have had other requests for access to data or individuals, which have been assessed by the team, illustrating the utility and relevance of the data being collected.

8.5 Future work

Having mentioned many areas that have created difficulties within Dogslife, I need to return to the positives. We have published new findings about the morphology and lifestyle of LRs in the UK (Pugh et al. [2015b]) and there were over 6,000 reports of illness in the first three and a half years of the project (Pugh et al. [2015a]). These illness reports are already being exploited in smaller case-control studies comparing high density SNP data such as the investigation of limber tail (Chapter 7). The case definitions were based on owner reports and the controls were chosen according to whether they had a similar amount of time in the project.

Given the level of under-reporting, we anticipated that some of the controls would have suffered from the conditions of interest and this was certainly the case for limber tail. If future genetic analyses based on online Dogslife reports find positive results, questionnaires will be required to definitively determine control status. Approximately 100 Dogslife dogs have been SNP chipped to date and, as more genotype data are generated, it will be feasible to speculatively run this type of analysis for a wide variety of illnesses.

It would also be interesting to investigate the morphology of the cohort. Can the heavier or lighter dogs be identified from their genotypes and with that type of information, can future generations of veterinarians advise owners about risk? We have found that chocolate coloured dogs are heavier than their yellow and black counterparts and hope to investigate that further. Attempts were made to identify yellow dogs with the chocolate mutation but none were found from an initial 25 dogs (Section 4.3.10). We have now targeted owners of yellow dogs that have pink or brown noses (indicative of the chocolate mutation) via the newsletter and hope to collect saliva samples from them for future investigations.

The height model detailed in Section 4.3.9 already includes a growth rate parameter. It was largely overlooked in the discussion because the aim of the modelling process was to deal with unit errors in data entry and ascertain a full height for the dogs. The model was not specified to best identify growth rates and would be particularly poor at younger ages. Nevertheless, the idea of determining growth rates for the dogs is intriguing and could borrow from a wealth of work in the livestock sector. If the data supported finding reliable growth rates for individual dogs then assessing whether this was associated with future musculoskeletal problems would be worth pursuing.

At this relatively early stage, a wealth of information has already been collected regarding lifestyle and lameness. A major avenue for future investigation will be to pick apart temporal associations regarding morphology, early exercise and subsequent diagnoses of musculoskeletal problems such as hip and elbow dysplasia. As the dogs continue to age, there will be potential to try to determine whether we can predict the onset of secondary osteoarthritis.

To date, geographical analyses have been based on latitude, longitude, postcode area and country boundaries within the UK. The analyses were limited by the way that the different countries collect information which meant that there was no consistent UK-wide definition of the urban-rural spectrum. As such, it was not possible to use a UK-wide measure for population density. Indeed, there is much argument about how population density should be defined. Future work could involve collaborations with geographers or experts in mapping who might facilitate this type of analysis. As Chapter 5 demonstrated, geographical variation was a factor in rates of gastrointestinal illness and better defining geographical factors for all UK postcodes would make it possible to include all Dogslife dogs in analyses.

Chapter 5 included a multivariable analysis of diarrhoea and vomiting rates, considering demographic characteristics, geographic variation and month of the year. Unfortunately the 'best' multivariable model including month of the year could not be resolved for vomiting and, rather than wait for more data as was previously suggested (Section 5.3.4), the modelling might be retried using new methodology. In June 2015, Torman and Camey [2015] published work which demonstrated the utility of applying Bayesian methods to this type of data. Their work has highlighted the possibilities available beyond traditional epidemiological approaches and is an example of how methodology will change over the period of data collection. Dogslife have a wealth of data and the large number of contributors make it an ideal data-set to try out multiple investigative techniques.

8.6 Future work without constraints

The first aim if Dogslife operated in a world without financial constraints would be to guarantee that data collection ran until at least 2025 (15 years after the oldest dogs were born). This would give complete life time data for thousands of dogs which would be a unique resource for investigation. We might better maintain owner retention by paying for yearly clinical check-ups for their dogs and thereby also gain definitive veterinary diagnoses of ongoing conditions. Ideally, dogs with conditions of interest would also be seen by Dogslife veterinarians to reduce inter-observer variability. As the limber tail investigation demonstrated (Section 7.3.2), it was hard to develop clear case definitions based on owner interpretation and reporting of signs.

If yearly check-ups became part of the study protocol, blood, urine and faecal samples could be collected by the clinician. Investigations of the ageing of the microbiome would be greatly facilitated by having repeated faecal samples from thousands of dogs. The GR study in the USA report that sample collection will be part of their study protocol (Guy et al. [2015]) and it will be interesting to see whether they attempt this type of analysis. Ideally, Dogslife results would be compared with those from the GR study because the USA has a different regulatory framework regarding anti-microbial use and, beyond breed differences, it may be possible to identify differences between gut microbiota in the two study populations.

Even within the UK, it would be a positive step to widen Dogslife out to more breeds, or specifically to dogs from no particular breed. Choosing dogs from no particular breed would create difficulties because the population of people who own non-pedigree dogs is not well characterised in the UK (Section 2.2) but this is an issue which has presumably been resolved within the Bristol Cats Study which makes no distinction between cat breeds for recruitment purposes. A comparison between the lives of LRs and non-pedigree dogs would be fascinating and potentially add to the wealth of work assessing health problems in pedigree dog populations which was recently reviewed by Farrell et al. [2015].

With unlimited money someone on the team would be trained to be able to make amendments to the website and database that can currently only be made by external contractors. This person would be in charge of maintaining the database, ensuring that issues such as the random duplications were identified quickly and resolved before they became a major problem. Above anything else, having this sort of expertise immediately available would greatly reduce the amount of time required to undertake analyses.

With someone on the team dedicated to the technical side of the project, it would be possible to keep Dogslife more up to date with how people access the internet. A questionnaire app might be developed for multiple platforms to enable owners to keep up with their Dogslife record wherever they might be. This could include a facility to request photographs of specific conditions such as limber tail for automatic upload by owners or clinicians.

There was some indication that recall decay would start to have more of an impact after the dogs reached one year and owners started answering the questionnaire quarterly rather than monthly. Ideally, the validation process would be repeated with another series of visits and collection of more veterinary records from this older group of dog so that the difference could be quantified. This might also facilitate a better understanding of which types of veterinary-visiting signs were more likely to be under-reported. Similarly, if an illness diary were sent to a subset of the owners, we might better be able to capture under-reporting of all signs, including those that did not precipitate a veterinary visit.

There is some concern that data collected through citizen science might be affected by the biases of the citizens involved (Nature Editors [2015]) so it would be of real interest to interview a number of contributors and non-contributors to try to determine what prevented people from joining Dogslife. Given that LRs are such a numerous breed, it is to be hoped that if any members of Dogslife are giving deliberately skewed information, they are overwhelmed by the wealth of other data. Nevertheless, identifying how the contributing cohort differ from the wider LR-owning public would be the final validation step.

As Chapter 6 demonstrated, the exercise section of the Dogslife questionnaire relates to a small part of each dog's day. It would be interesting to try to fill in the gaps by developing a different type of questionnaire. Dogs in retired households spent the least amount of time being sedentary and one might hypothesise that this relates to the amount of time the dogs spend alone. A retired owner might have more time with their dog and the dog would therefore spend less time asleep or resting. The presence of another dog seemed to have no effect so we might ask owners how long on average each day their dog is left with no people for company. Ideally this type of question would again be validated by sending GPS devices and accelerometers to a subset of owners. They would be permanently attached and downloaded automatically via a smart phone. In order to maximise the potential of that approach, subjective measures such as how active each owner perceives their dog to be might also be assessed.

Age was also highlighted as important with regard to accelerometer readings so wider work whereby individual dogs are sent accelerometers repeatedly throughout their lives could also yield interesting findings. Owners are currently advised to limit the activity of their young dogs but such advice would be redundant if the young dogs are independently much more active than their older counterparts. To my knowledge, nobody has addressed activity levels in young dogs, presumably because of the complexities of working with a variable that is assumed to change with age. New recruits to the Dogslife cohort could become part of a specific activity study but we are currently hampered by the expense of such an approach. The accelerometers used in the exercise chapter cost £300 each and the base-station used to download the data was another £1,000. Recent work by Yashari et al. [2015] has validated a cheaper smart phone based approach to collecting activity data but their devices had a significantly shorter battery life and still cost $79 each.

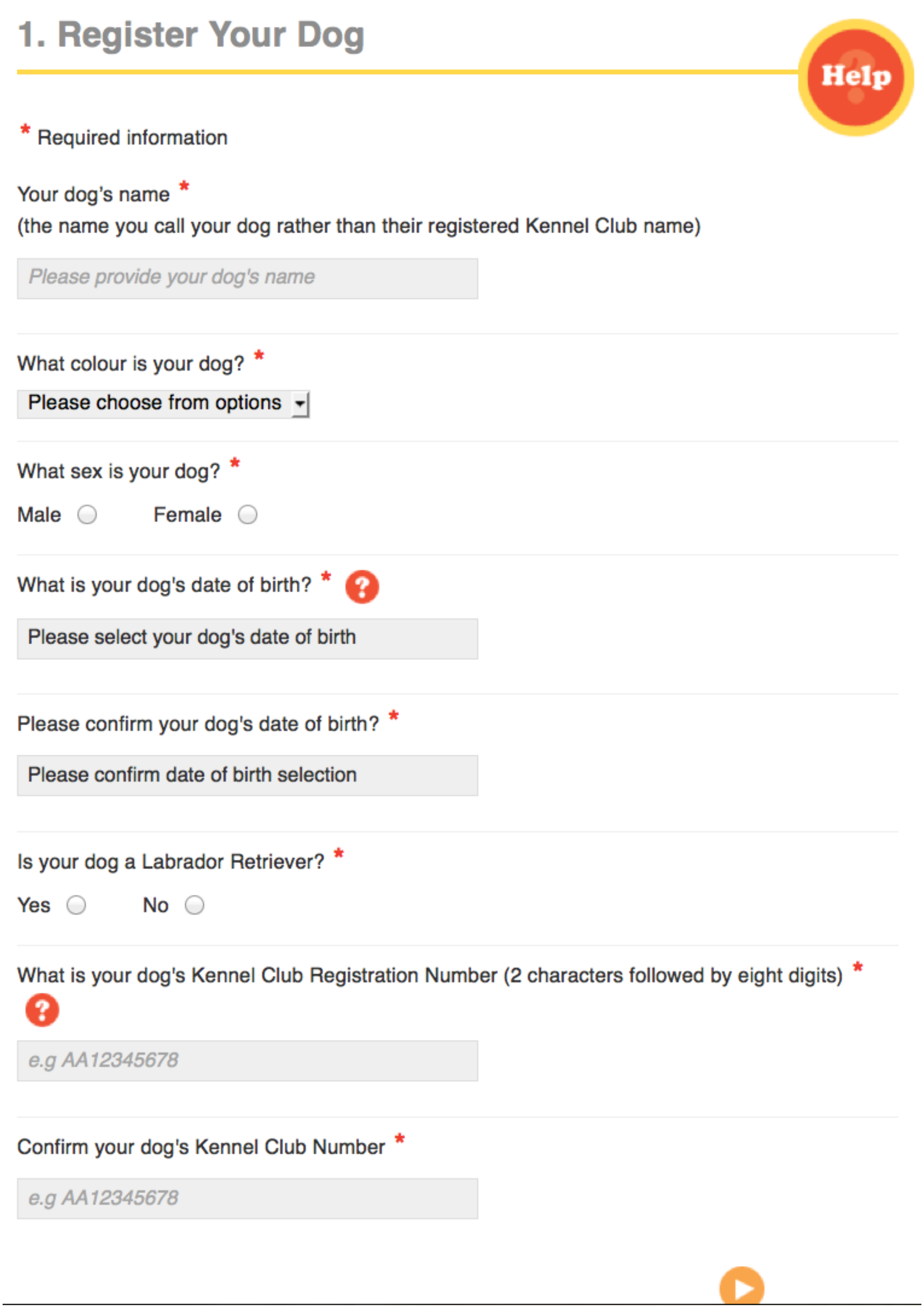
If it were possible to fund widespread use of accelerometers or similar devices amongst the Dogslife cohort from a young age, a number of those dogs would subsequently be diagnosed with hip and elbow dysplasia. Undoubtedly, many would also subsequently develop obesity. Detailed analysis of activity in an ageing cohort could contribute fascinating new information about the relationship between exercise, morphology and developmental musculoskeletal problems.

The Dogslife cohort have so much potential to help answer questions about canine health. With more resources, specific omissions such as collection of DNA from the entire cohort could be remedied. It would be possible to bypass using a SNP chip which was been designed to apply to all breeds and move immediately to next generation sequencing for all participating dogs. Beyond that, there is immense scope to extend the study and make the most of the efforts of a committed group of citizen scientists.

8.7 Conclusion

Working on the Dogslife project has been a unique opportunity to begin a new chapter in the story of canine health, building the foundations of a new era of holistic, evidence-based canine medicine. I hope that this thesis and the associated publications will be invaluable for future investigators, particularly those who will work on the Dogslife project as it moves into its sixth year and beyond.

Appendix 1: Registration



1. Register Your Dog

Help

* Required information

Your dog's name *
(the name you call your dog rather than their registered Kennel Club name)

Please provide your dog's name

What colour is your dog? *

Please choose from options

What sex is your dog? *

Male Female

What is your dog's date of birth? *

Please select your dog's date of birth

Please confirm your dog's date of birth? *

Please confirm date of birth selection

Is your dog a Labrador Retriever? *

Yes No

What is your dog's Kennel Club Registration Number (2 characters followed by eight digits) *

e.g AA12345678

Confirm your dog's Kennel Club Number *

e.g AA12345678

Figure A1: Dog registration

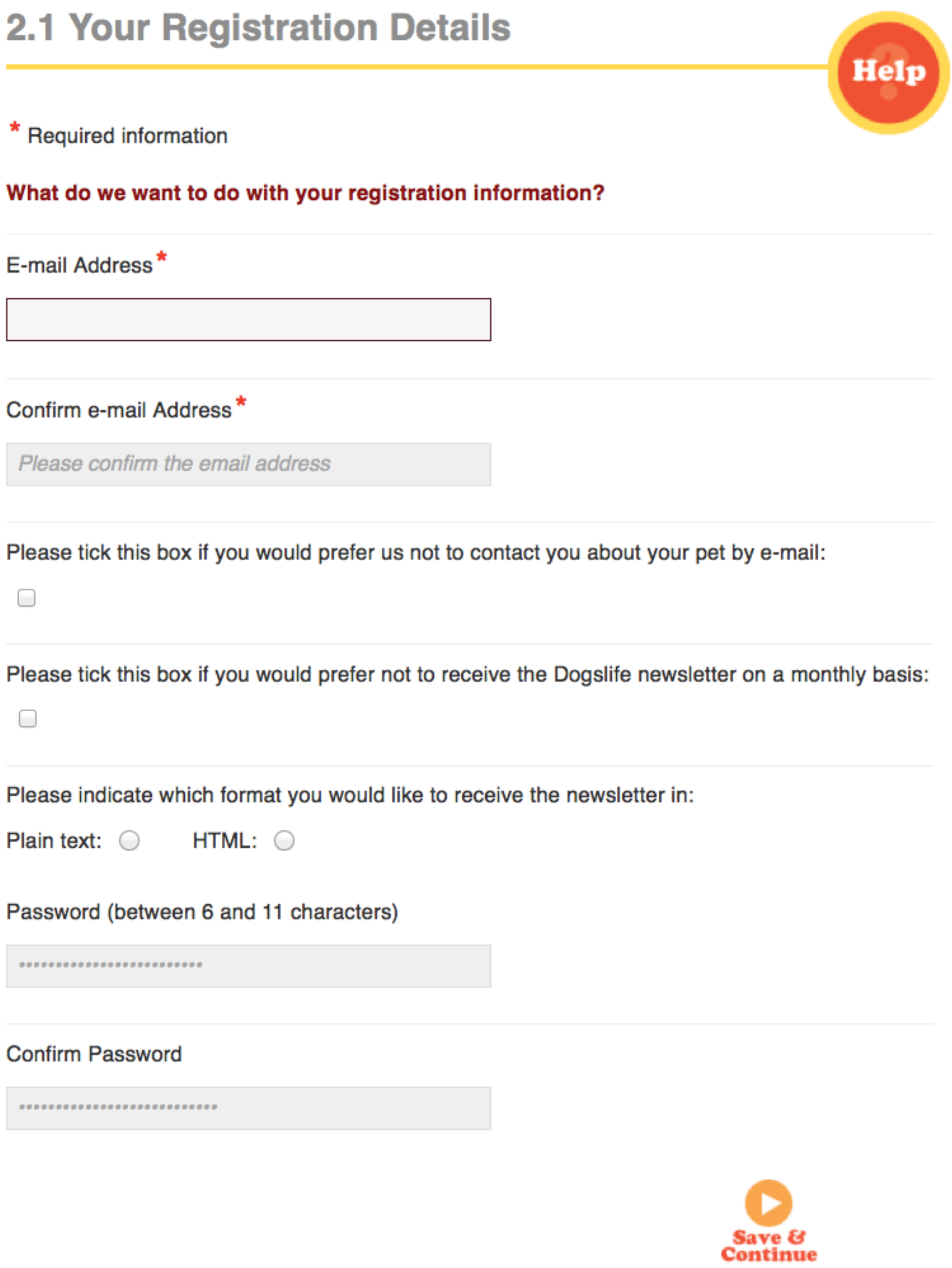
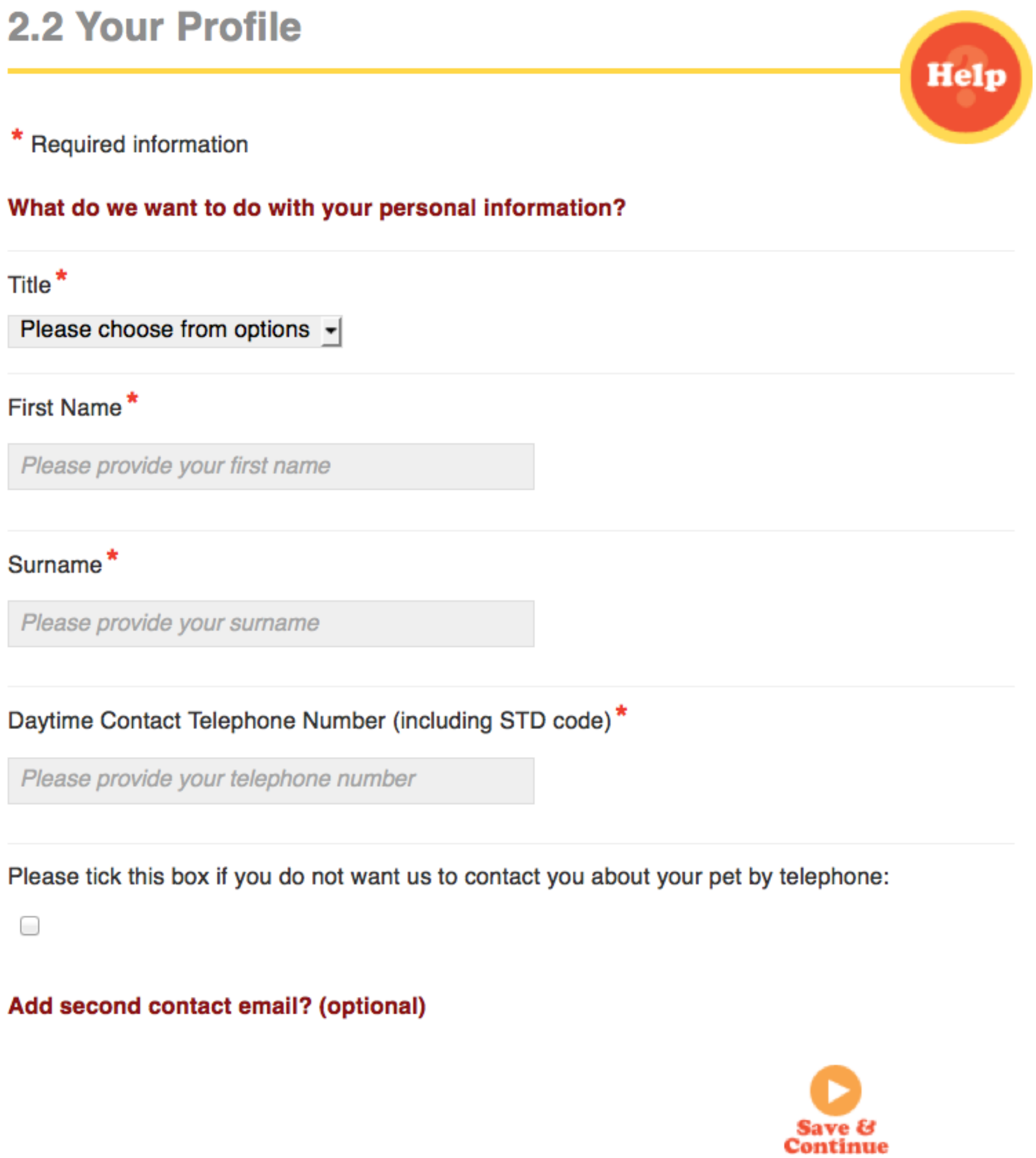


Figure A2: Owner registration and email details



2.2 Your Profile

Help

* Required information

What do we want to do with your personal information?

Title *

Please choose from options

First Name *

Please provide your first name

Surname *

Please provide your surname

Daytime Contact Telephone Number (including STD code) *

Please provide your telephone number

Please tick this box if you do not want us to contact you about your pet by telephone:

Add second contact email? (optional)

Save & Continue

Figure A3: Owner name and telephone details

Figure A4: Household information

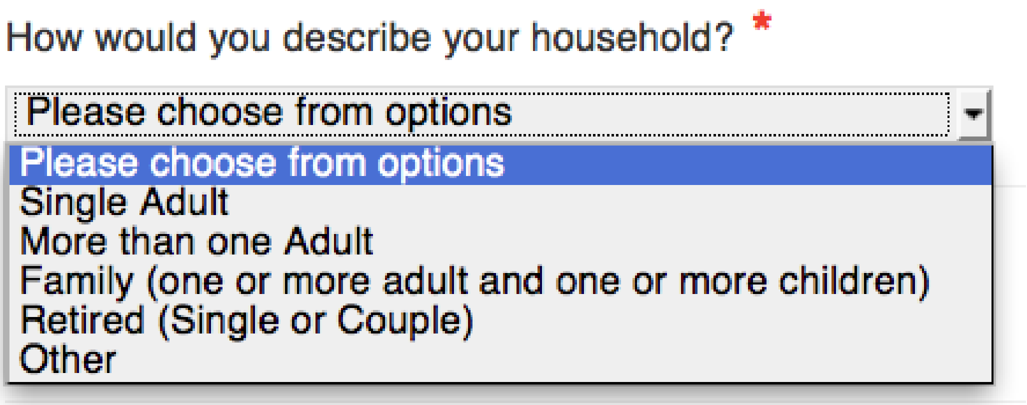


Figure A5: Household type options

Appendix 2: Dogslife Questionnaire

A2.1 Introduction

The Dogslife questionnaire below is broken into sections for each of the separate pages on the website: Pet profile A2.2; Sleeping & bathing A2.3; Exercise A2.4; Feeding A2.5; Preventative health care A2.6; Illnesses A2.7. Doggie* has been used as the name of a participating dog. Whenever 'Other' is selected, a free-text box appears where the owner might write their answer. They are not obliged to write anything in the box to continue through the questionnaire. After each page, the owner must click a button to save and continue on through the questionnaire. With the exception of dog weight, they cannot continue where answers to main questions have been omitted. Additional questions that appear after an owner has ticked Yes or No may be optional and may therefore be left blank (these are marked below). The first time the owner answers the questionnaire, the phrasing 'since you last visited the site' is replaced with 'in the last four weeks'. The only exception is for vaccinations whereby the question simply starts as 'Has Doggie* been vaccinated?'

A2.2 Pet Profile

Has Doggie* been neutered since you last visited the site? [Yes/No]
If 'Yes', **When was Doggie* neutered? If you don't know the exact date, please enter an approximate date.** [Pop-up calendar]

Why did you choose to have Doggie* neutered?

- Recommended by vet
- Recommended by breeder
- I do not want to breed from my dog
- Behavioural reasons
- Other

If 'No', **Do you plan to breed from Doggie*?** [Yes/No/Undecided]
How tall is Doggie*? [Number][Centimetres/Inches]
How much does Doggie* weigh? [Number][Kgs/Pounds]
Do you own Doggie* primarily as a:

- Household pet
- Working dog (e.g. Gundog)
- Assistance dog (e.g. Guide Dog)
- Other

(This question is only asked once, when the owner first fills in the questionnaire.)

A2.3 Sleeping & Bathing

Where does Doggie* sleep at night?

- Alone in a room in house
- In a room shared with a person
- In a room shared with a pet and a person
- Outside
- Other

Has Doggie* been bathed since you last visited the site? [Yes/No]
If 'Yes', **On average, how often do you bathe Doggie**

- Daily
- Twice weekly
- Weekly
- Fortnightly

– Monthly
– Less than monthly
– Never

What do you use to bathe Doggie*?

– Water alone
– Dog shampoo
– Human shampoo
– Other

Have you registered Doggie* with a veterinary practice since you last visited the site? [Yes/No]
If 'No', then ask again at revisit.
Have you had Doggie* insured since you last visited the site? [Yes/No]
If 'Yes', then don't ask again for 1 year.
If 'No', then ask at revisits until answer changes to 'Yes'.

A2.4 Exericse

The owner is asked to give an answer for weekdays and weekend days. **On average, in the last week for how long does Doggie* do the following exercise(s) EACH DAY?**

– Walking on the lead
– Running on the lead
– Walking/running off the lead
– Exercise involving fetching, chasing and retrieving
– Obedience training
– Other playing activity (including dogs playing together)

They are asked to choose between the following:

– None
– 1-5 minutes

- 5-15 minutes
- 15-30 minutes
- 30 minutes - 1 hour
- 1-2 hours
- Over 2 hours

Is the quantity of exercise Doggie receives each day:

- Restricted due to your own time constraints
- Restricted due to your own exercise ability
- Restricted because of where you live
- Restricted because Doggie* has a problem
- As recommended by my dog breeder/my own experience
- Unrestricted (you give as much exercise as you think your dog should have)

A2.5 Feeding

At present, how many times a day do you feed Doggie*?

- Once daily
- Twice daily
- Three times daily
- More than three times daily
- Throughout the day

When do you feed Doggie*?

- In the morning
- In the evening
- In the morning and evening
- In the morning, lunchtime and evening
- Throughout the day
- Multiple times throughout the day

What types of food do you give to Doggie*?
(Multiple selections permitted†)

- Dried food[1]
- Tinned food[1]
- A mixture of dried and tinned dog food[1]
- Home prepared food[2]
- Other[2]

If foodtype[1], **Which [foodtype] do you feed Doggie*?** (Optional question) [Aldi/Asda/Bakers/Butchers/Brunos/Cesar/Chappie/High Life/Hills/Iams/ James Wellbeloved/Lidl/Pedigree/Royal Canine/Sainsburys/Tesco/Wagg/Wilson/ Winalot/Other]
If foodtype[2], **Please detail the food content.** (Optional question) [Free text box]
How much foodtype do you feed Doggie* each day? [Number][Grams/Ounces] (Asked for each selection†)
Does Doggie* also receive "titbits"? For example, anything else your dogs east such as food off your plate, training treats, chews etc? [Yes/No]
What does Doggie* drink most days? [Water/Other]

A2.6 Routine Health Care

Has Doggie* been vaccinated since you last visited the site? [Yes/No]
If 'Yes', **When was Doggie* vaccinated?** [Pop-up calendar]
Which diseases was Doggie* vaccinated against? (Optional question) Additional information is available by clicking *i* which states: **Routine vaccination protects agains distemper, canine hepatitis, parvovirus or leptospirosis**

- Routine vaccinations
- Kennel cough
- Rabies
- Other

Has Doggie* been wormed since you last visited the site? [Yes/No]
If 'Yes', **When was Doggie* last wormed?** [Pop-up calendar]

What is the name of the product you use? (Optional question)
[Droncit/Drontal/Granofen/Milbemax/Panacur/Piperazine/Plerion5/
Vert-X Treats for Dogs/Wormazole/Other]
Have you used any products to prevent or treat fleas or ticks since you last visited the site? [Yes/No]
If 'Yes', **What type of flea control product do you use?** (Optional question)
[Collar/Oral Medication/Drops or "Spot-on" for the dog's coat/
Spray for the dog's coat/Spray for the house/Other]
Have you travelled abroad (out of the UK) with Doggie* since you last visited the site? [Yes/No]
If 'Yes', **Please provide details of date and location.** (Optional question)
[Free text box]

A2.7 Illness

Has Doggie* had any of the following problems since you last visited the site?

- Vomiting [Yes/No]
- Diarrhoea [Yes/No]
- Coughing [Yes/No]
- Scratching themselves [Yes/No]
- Licking or chewing themselves [Yes/No]
- Limping or lameness [Yes/No]

Did Doggie have any other illnesses or problems? [Yes/No]
If 'Yes', **What was the problem?** (Optional question) [Free text box]
All further questions are optional.
If 'Yes' to any, for each selection, **Approximately when did the [insert illness name] start?** [Pop-up calendar]
Approximately when did the [insert illness name] get better? If your dog is not better yet, click the box below.
Approximately how often did the [insert illness name] happen?
[Once/Continuous/Every hour/Every 2 hours/Every 6 hours/
Every 12 hours/Once a day/Once every 2 days/Once every 3 days/

Once weekly/Once every two weeks/Once a month]

Did you take Doggie* to the vet for the [insert illness name]? [Yes/No]

If 'No' (no veterinary visit), **Do you know why Doggie* developed this problem?**

If 'Yes' (veterinary visit), **Approximately when did you visit the vet?** [Pop-up calendar]

Did the vet fill in your dog's Dogslife Veterinary Health Report? [Yes/No]

If 'Yes' (health report completed), **From the information on the Dogslife Veterinary Health Report.**

What was the primary presenting complaint? [Free text box]

What is the primary Diagnosis? [Free text box]

What is/are the treatment/s for the primary condition? [Medical/ Surgical/Both/None/Other]

Treatment 1 [Free text box]

Treatment 2 [Free text box]

Treatment 3 [Free text box]

If 'Yes', **What was the secondary presenting complaint (only fill in if applicable)?** [Free text box]

What is the secondary Diagnosis? [Free text box]

What is/are the treatment/s for the secondary condition? [Medical/ Surgical/Both/None/Other]

Treatment 1 [Free text box]

Treatment 2 [Free text box]

Treatment 3 [Free text box]

If 'No' (health report not completed), **Download the Dogslife Veterinary Health Report now for the next time you visit your vet.**

Do you know what the diagnosis was? [Free text box] **What is/are the treatment/s?** [Medical/Surgical/Both/None/Other]

Treatment 1 [Free text box]

Treatment 2 [Free text box]

In addition, if 'Yes' to Limping or lameness, **Which limb was affected?** [Left front leg/Right front leg/Left back leg/Right back leg] (Multiple selections permitted)

Appendix 3: Visit Questionnaire

Visit questionnaire

Introduction

- Thank you
- I don't know any of the data you have entered - can't influence answers
- Very similar to online questionnaire
- Have answers changed and if so, how
- Can questions be improved

The Obvious

1. **What sex is Name?** ☐ M ☐ F
2. **What colour is Name?** ☐ B ☐ Y ☐ C

Demographics

Person name and title don't need validation. Will need to confirm postcode before visiting.

3. **How would you characterise your household?**
   - ☐ Single adult
   - ☐ More than one adult
   - ☐ Family (adult and one or more children)
   - ☐ Single or couple retired
   - ☐ Other: ______________________
4. **Does anyone in your household smoke?** ☐ Yes ☐ No
5. **Are there any other pets in the household?** ☐ Yes ☐ No
   - ☐ Dogs? (how many) ______________________
   - ☐ Cats? (how many) ______________________
   - ☐ Others? (what and how many)

   ____

   ____

   ____

Pet Profile

6. **What is Name's purpose?**
   - ☐ Household
   - ☐ Working dog (ie. gun dog)
   - ☐ Assistance dog (ie. guide dog)
   - ☐ Other: ______________________
7. **Has Name been neutered?** ☐ Yes ☐ No

8. Do you plan to breed from Name? ☐ Yes ☐ No ☐ Don't know

9. Where does Name currently sleep at night?
- ☐ Alone in room
- ☐ In a room shared with a member of the family
- ☐ Outside
- ☐ Other: ____________________

10. Has this changed in the time that you've owned Name?
- ☐ Yes
- ☐ No

If yes then go to next question If no then skip next question

11. Where did Name used to sleep at night?
- ☐ Alone in room
- ☐ In a room shared with a member of the family
- ☐ Outside
- ☐ Other: ____________________
- ☐ Comment: ________________________________

A

12. Do you...and if so, how often? (tick and circle)
- ☐ Swim Name in the sea/ a lake/ a river:
  - Daily
  - Twice weekly
  - Weekly
  - Fortnightly
  - Monthly
  - Less than monthly
- ☐ Hose Name off:
  - Daily
  - Twice weekly
  - Weekly
  - Fortnightly
  - Monthly
  - Less than monthly
- ☐ Put Name in the shower:
  - Daily
  - Twice weekly
  - Weekly
  - Fortnightly
  - Monthly
  - Less than monthly
- ☐ Put Name in the bath:
  - Daily
  - Twice weekly
  - Weekly
  - Fortnightly

- Monthly
- Less than monthly

13. How often do you bathe name?

- ☐
- ☐
- ☐
- ☐ Other: ______________________

14. What do you use to bathe name?

- ☐ Water alone
- ☐ Dog shampoo
- ☐ Human shampoo
- ☐ Other: ______________________

15. Would you consider the following to be bathing?

- ☐ Swimming Name in the sea/ a lake/ a river
- ☐ Hosing Name off
- ☐ Putting Name in the shower
- ☐ Putting Name in the bath

Comment ______________

16. Is Name registered at a veterinary practice? ☐ Yes ☐ No

17. Would you be happy for us to contact your veterinary practice to ask for Name's records?
☐ Yes ☐ No

Name/Address/Phone

____

____

____

____

____

18. Is Name insured? ☐ Yes ☐ No

Exercise

		Lead Walk	Lead Run	Off Lead
Weekday	None			
	1-5 minutes			
	5-15 minutes			
	15-30 minutes			
	30-60 minutes			
	1-2 hours			
	Over 2 hours			
Weekend	None			
	1-5 minutes			
	5-15 minutes			
	15-30 minutes			
	30-60 minutes			
	1-2 hours			
	Over 2 hours			

		Fetch, chase or retrieve	Obedience Training	Other play
Weekday				
	None			
	1-5 minutes			
	5-15 minutes			
	15-30 minutes			
	30-60 minutes			
	1-2 hours			
	Over 2 hours			
Weekend				
	None			
	1-5 minutes			
	5-15 minutes			
	15-30 minutes			
	30-60 minutes			
	1-2 hours			
	Over 2 hours			

19. Is the quantity of exercise Dog receives each day:

- ☐ Restricted due to your own time constraints
- ☐ Restricted due to your own exercise ability
- ☐ Restricted because of where you live
- ☐ Restricted because Dog has a problem
- ☐ As recommended by my dog breeder / my own experience
- ☐ Unrestricted (you give as much exercise as you think your dog should have)

Diet

20. How many times per day do you feed Name at the moment?

- ☐ Once
- ☐ Twice
- ☐ More than twice
- ☐ Throughout the day (don't include titbits or rewards)

21. When do you feed Name?

- ☐ Morning
- ☐ Evening
- ☐ Morning & evening
- ☐ Throughout the day

22. What types of food do you give Name?

- ☐ Dried commercial
- ☐ Tinned commercial
- ☐ Mixture of dried and tinned dog food
- ☐ Home prepared food
- ☐ Other: ____________________

23. What brand(s) do you feed Name?

____

____

____

24. How much of each food to you feed Name each day? (grams)

____

____

25. Does Name also receive titbits or left-over food from your own meals/snacks? ☐ Yes ☐ No

26. Ask if I can weigh the food

. Food type and weight:

____

____

____

Routine Health Care

27. Has Name been given any vaccinations? ☐ Yes ☐ No ☐ Don't know

☐ Routine Vaccinations
☐ Kennel Cough
☐ Rabies

28. Do you have a vaccination card and if so, may I see it? ☐ Yes ☐ No

29. Did I see vaccination card? ☐ Yes ☐ No

Details:

____

____

____

____

____

30. Do you worm Name? ☐ Yes ☐ No

____

If yes then go to next question If no then skip next question

31. How do you apply the wormer and what is it's name? (If more than one brand then include all)

☐ Spot-on
☐ Oral
☐ Seen?
Brand(s) ______________

32. Do you treat Name for fleas? ☐ Yes ☐ No

____

If yes then go to next question If no then skip next question

33. How do you apply the flea treatment and what is it's name? (If more than one brand then include all)

☐ Spot-on
☐ Oral

- ☐ Powder
- ☐ Seen?

Brand(s) ____________

Illness

34. Has Name been ill since you got him/her? ☐ Yes ☐ No

____

Ask the following irrespective of the answer above.

35. Has Name had any of the following:

- ☐ Vomiting - Vet Visited? ☐ Yes ☐ No
- ☐ Diarrhoea - Vet Visited? ☐ Yes ☐ No
- ☐ Coughing - Vet Visited? ☐ Yes ☐ No
- ☐ Scratching themself - Vet Visited? ☐ Yes ☐ No
- ☐ Licking or chewing themself - Vet Visited? ☐ Yes ☐ No
- ☐ Lameness or Limping - Vet Visited? ☐ Yes ☐ No
- ☐ Other ____________ Vet Visited ☐ Yes ☐ No

Comments:

____

____

____

____

Measurements

36. Did you view the video about measuring height? ☐ Yes ☐ No

37. Where do you measure height to on Name? ____________

38. Height measured? ☐ Yes ☐ No

____

39. Weight measured? ☐ Yes ☐ No

____

40. Vet permission sought? ☐ Yes ☐ No

Comments:

____

____

____

____

____

Appendix 4: Dogslife Health Report

Dogslife™ Veterinary Health Report

To be completed by the Vet.

Dogslife is an epidemiological study monitoring the health of Kennel Club registered Labrador Retrievers through a web-based data collection system. We would be grateful if you could summarise the findings of today's **NON-ROUTINE** veterinary presentation (i.e. not vaccination, flea treatment or worming) so that **this information can be recorded into the Dogslife record for the patient BY THE OWNER next time they visit the website.**
All the information you provide is anonymous. For further details visit **www.dogslife.ac.uk**

NAME:

DATE OF BIRTH:

PRIMARY PRESENTING COMPLAINT (e.g. vomiting / left hind limb lameness) ☐

PRIMARY DIAGNOSIS (e.g. acute gastritis / hip dysplasia)? ☐

or DIAGNOSIS NOT MADE (no diagnosis) ☐ **DIAGNOSIS UNCHANGED** ☐

WHAT IS / ARE THE TREATMENT(S) FOR THE PRIMARY CONDITION?

Medical ☐ **Surgical** ☐ **Both** ☐ **None** ☐

Treatment 1 ☐

Treatment 2 ☐

Treatment 3 ☐

SECONDARY PRESENTING COMPLAINT (if applicable)**?** ☐

SECONDARY DIAGNOSIS (if applicable)**?** ☐

or DIAGNOSIS NOT MADE (no diagnosis) ☐ DIAGNOSIS UNCHANGED ☐

WHAT IS / ARE THE TREATMENT(S) FOR THE SECONDARY CONDITION?

Medical ☐ **Surgical** ☐ **Both** ☐ **None** ☐

Treatment 1 ☐

Treatment 2 ☐

Treatment 3 ☐

ADDITIONAL COMMENTS

☐

Thank you for your time and help.
Further information about the Dogslife project can be found at **www.dogslife.ac.uk.**
If you have any questions, please e-mail **info@dogslife.ac.uk**

Appendix 5: Height Model Checking

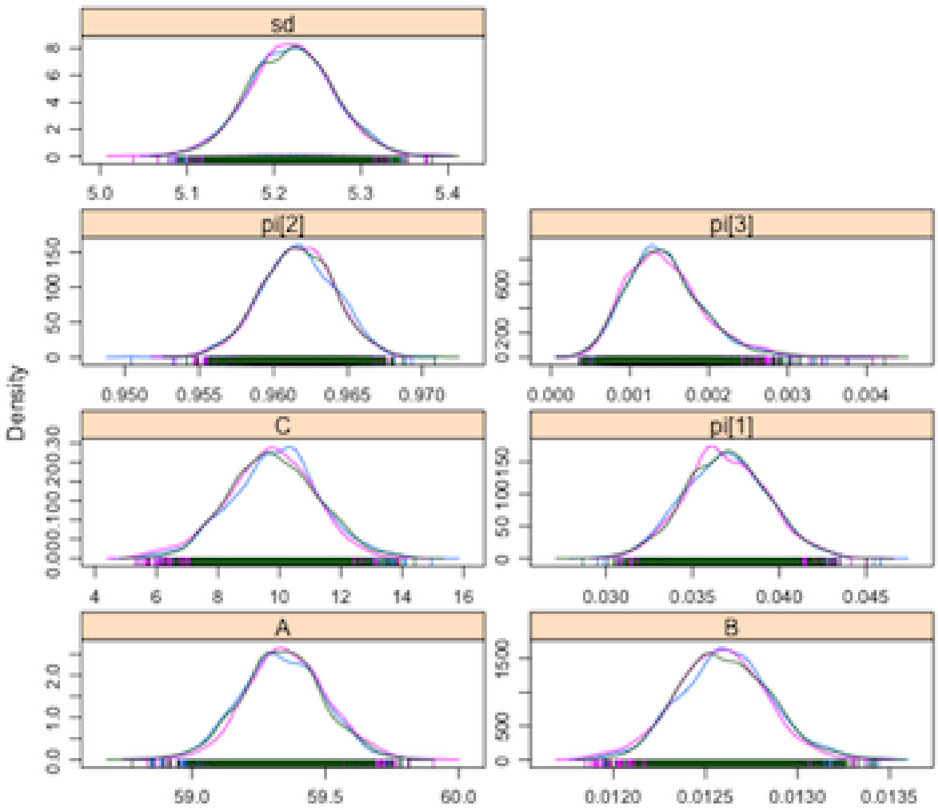


Figure A6: Density Plots Of Variables For Male Dogs

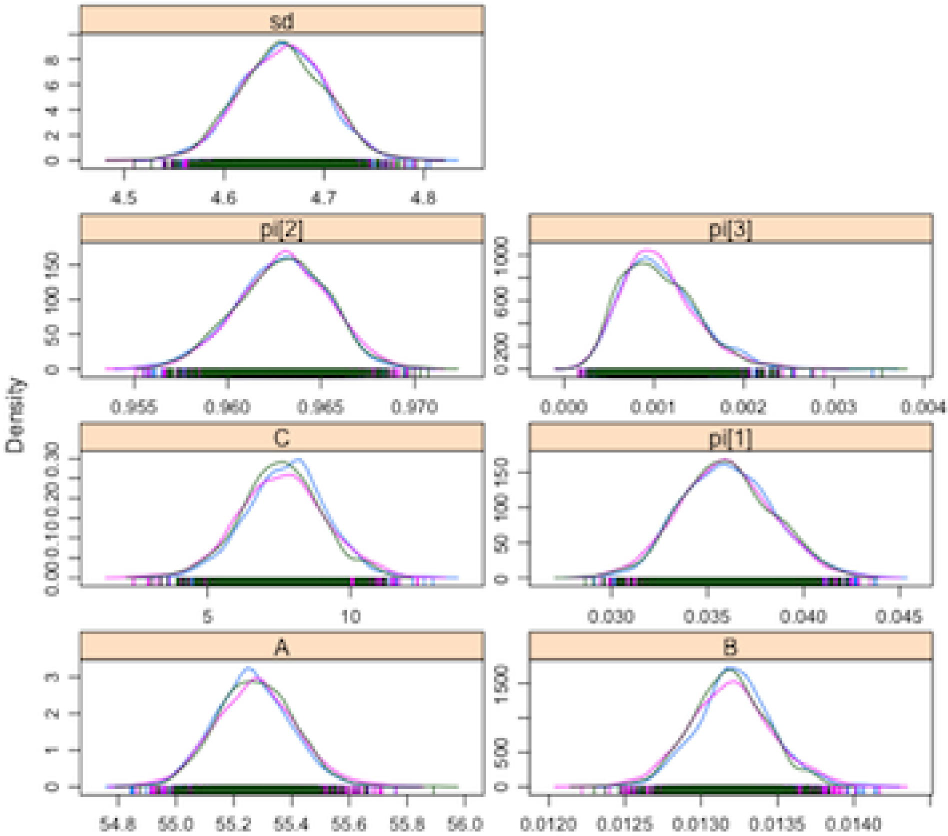


Figure A7: Density Plots Of Variables For Female Dogs

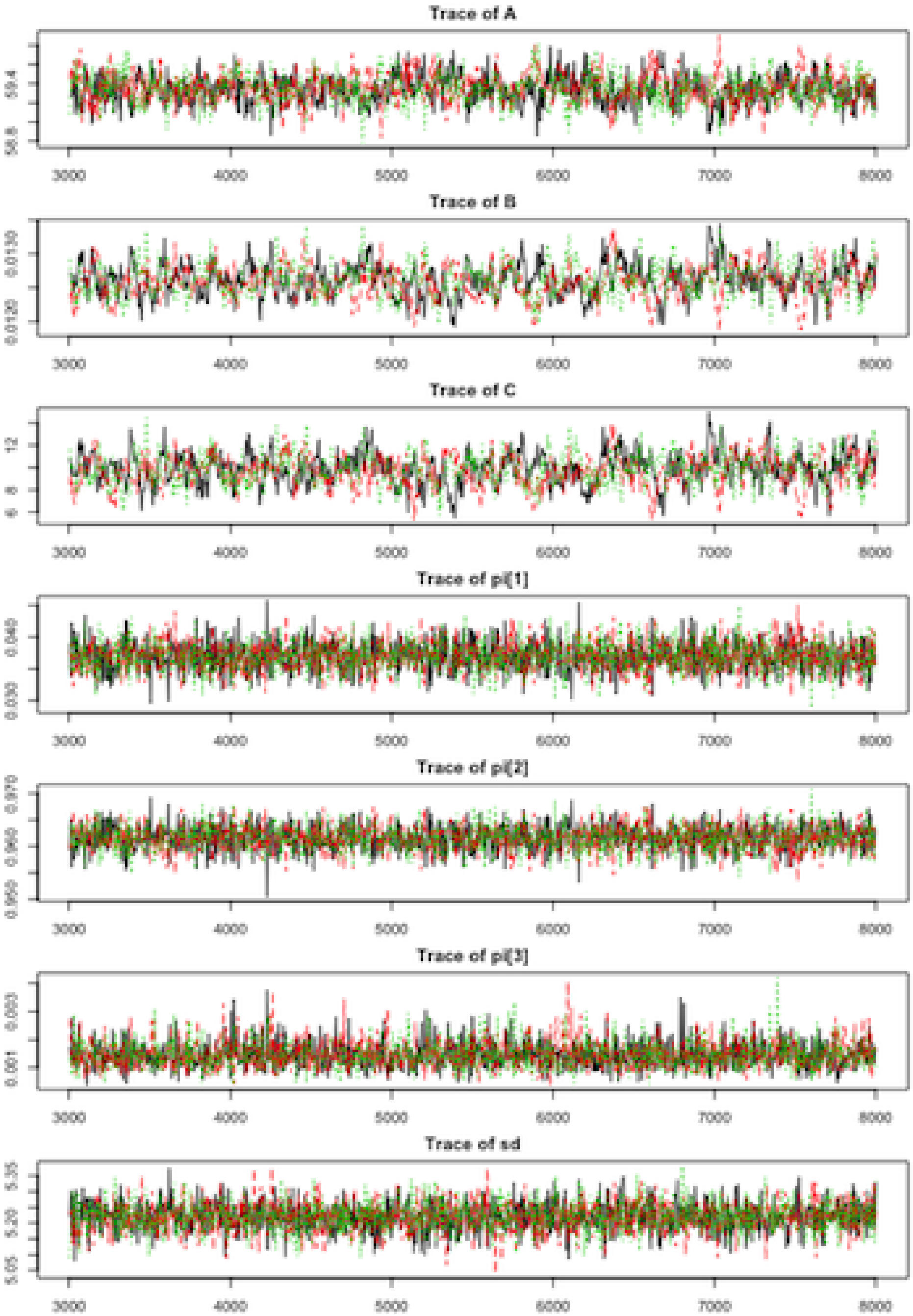


Figure A8: Mixing Trace Plots Of Variables For Male Dogs

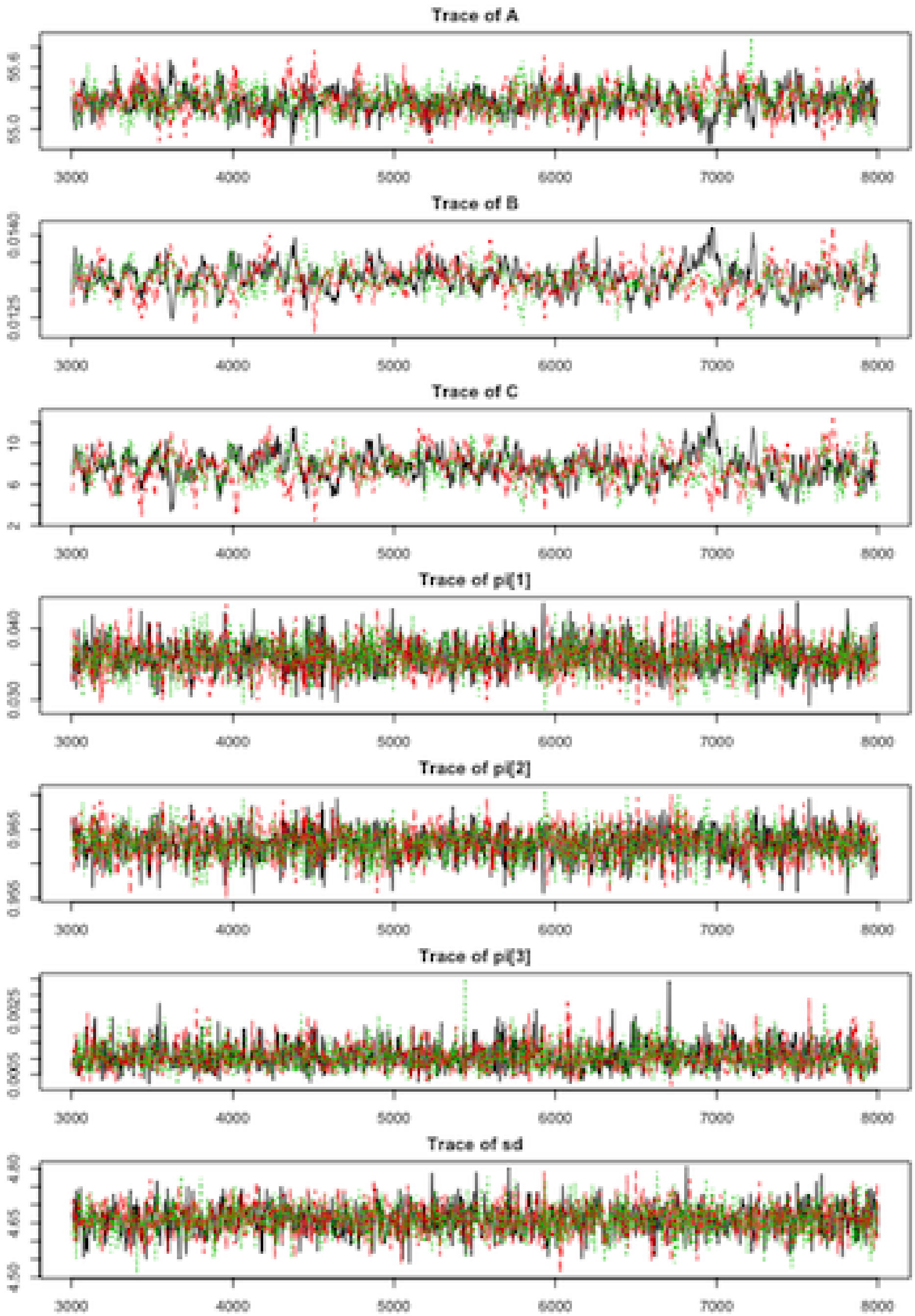


Figure A9: Mixing Trace Plots Of Variables For Female Dogs

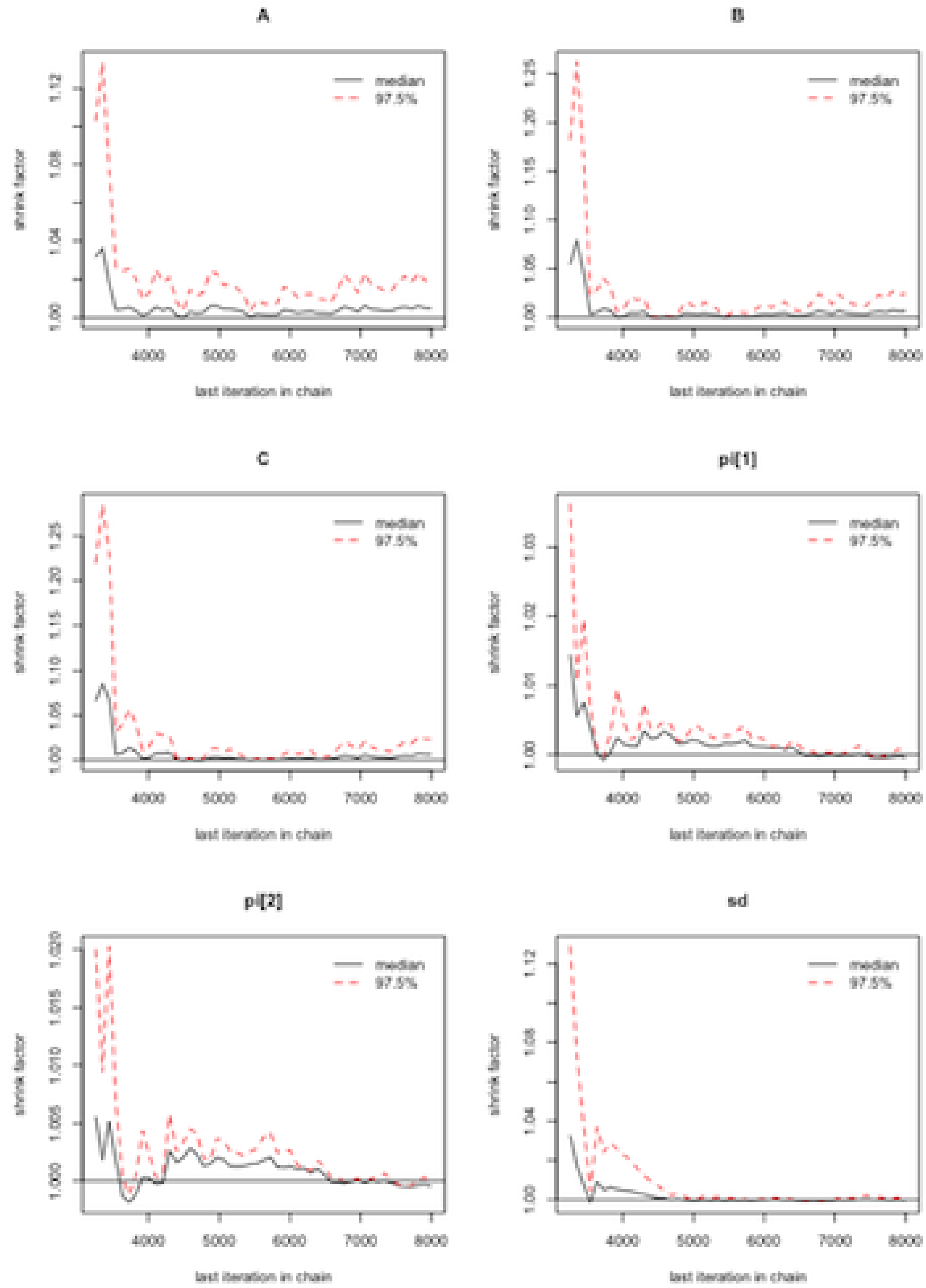


Figure A10: Gelman Plots To Show Mixing Of Variables For Male Dogs

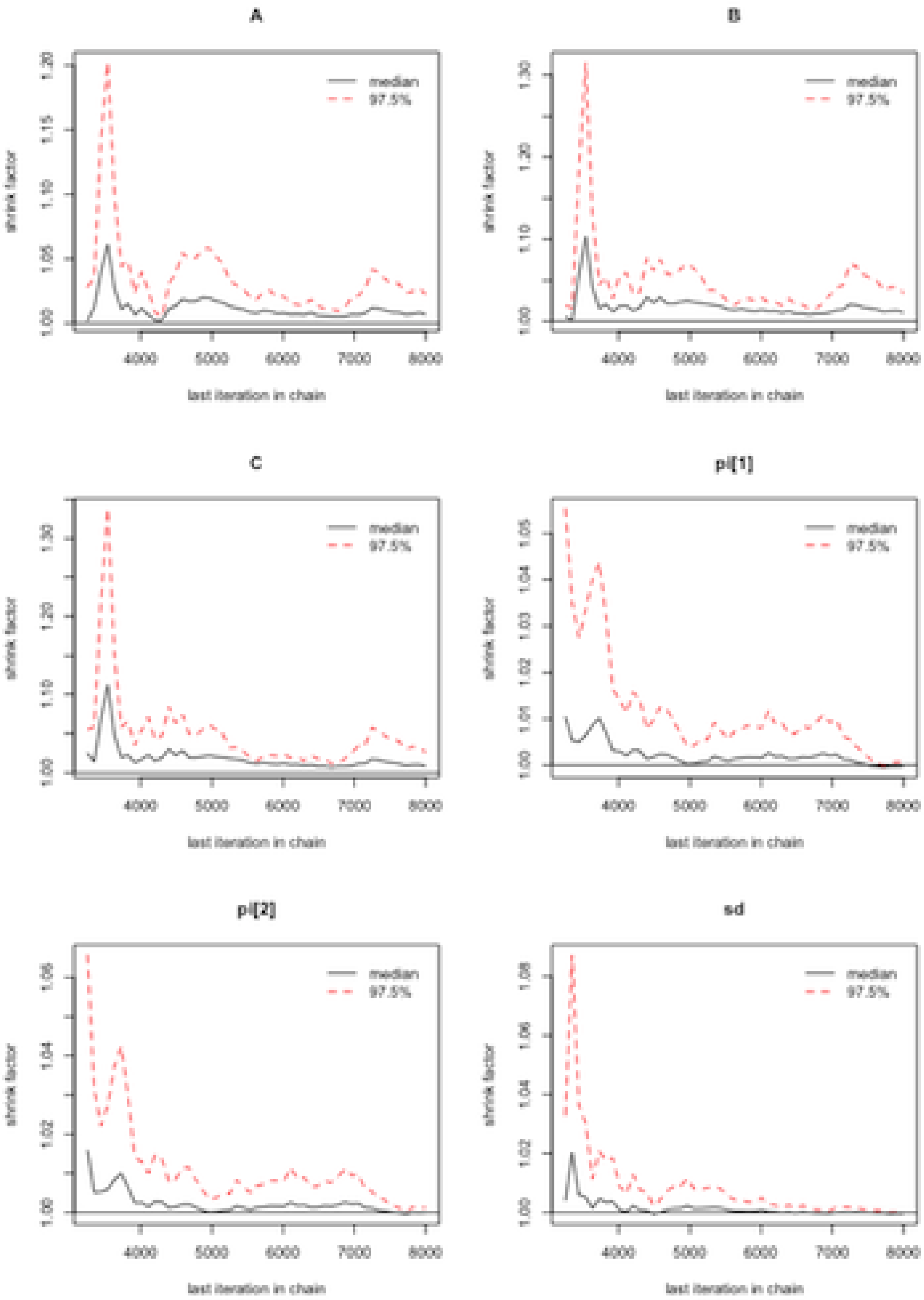


Figure A11: Gelman Plots To Show Mixing Of Variables For Female Dogs

Appendix 6: Illness Frequencies

Table A6.1: Frequency of All Presenting Signs

Sign	Frequency
Faecal appearance abnormal - diarrhoea	1622
Vomiting - other	1566
Pruritus	1003
Gait abnormality - lameness	819
Coughing	324
Ear (aural) abnormality	192
Ophthalmic (eye) abnormality	182
Presenting complaint not listed	104
Wound	82
Skin (cutaneous) abnormality - other	73
Mass/swelling - skin (cutaneous)	65
Discharge - ocular (eye)	57
Skin (cutaneous) abnormality - eruptions/hives/rash	46
Dietary indiscretion - other	43
Traumatic episode	41
Mass/swelling - other	39
Anal irritation	37
Urination abnormal - other	36
Vomiting - haematemesis	32
Dietary indiscretion - foreign body ingestion	22
Musculoskeletal injury	22

Continued on next page

Table A6.1: Frequency of All Presenting Signs (continued)

Sign	Frequency
Dental (tooth) abnormality	20
Discharge - vaginal	20
Seizure(s)	19
Urination abnormal - pollakiuria	19
Red eye	17
Faecal appearance abnormal - haematochezia	15
Mass/swelling - eye (ophthalmic)	15
Ophthalmic (eye) injury	15
Appetite decreased	13
Sneezing	13
Discharge - aural (ear)	12
Eyelid abnormality	12
Incontinence - urinary	12
Pyrexia/hyperthermia	11
Alopecia	10
Mass/swelling - oral (mouth)	10
Pain - ophthalmic (eye)	10
Lethargy	8
Mass/swelling - abdominal	8
Mass/swelling - ear (aural)	8
Oral (mouth) abnormality - other	7
Post-operative complication - wound related	7
Behavioural abnormality	6
Pain - musculoskeletal	6
Polyuria	6
Salivation - increased/drooling	6
Ear (aural) injury	5
Gagging/retching	5
Mass/swelling - musculoskeletal	5
Oral (mouth) injury	5
Polydipsia	5

Continued on next page

Table A6.1: Frequency of All Presenting Signs (continued)

Sign	Frequency
Urine colour abnormal	5
Abdominal distension	4
Collapse/syncopal episodes	4
Discharge - nasal (nose)	4
Intoxication (poisoning)/toxin exposure	4
Mentation altered - anxious/distressed	4
Obesity	4
Pain - aural (ear)	4
Tremors/shaking/trembling	4
Defaecation abnormal - other	3
Faecal appearance abnormal - other	3
Gait abnormality - ataxia	3
Pain - on eating/chewing	3
Pain - oral (mouth)	3
Panting	3
Paresis/paralysis	3
Urination abnormal - dysuria	3
Weakness	3
Weight loss	3
Cardiac (heart) abnormality - other	2
Defaecation abnormal - defaecatory tenesmus	2
Gait abnormality - other	2
Lymphadenomegaly	2
Mass/swelling - perianal	2
Pain - other	2
Stiffness	2
Testicular/scrotal abnormality	2
Urination abnormal - inappropriate urination	2
Urination abnormal - oliguria	2
Posture abnormal - other	2
Anorexia	1

Continued on next page

Table A6.1: Frequency of All Presenting Signs (continued)

Sign	Frequency
Bleeding - epistaxis	1
Blindness	1
Cataract	1
Colic	1
Discharge - penile	1
Dry eye	1
Failure to gain weight	1
Halitosis	1
Incontinence - faecal	1
Oral (mouth) abnormality - ulceration	1
Pain - abdominal	1
Penile/preputial abnormality - other	1
Penile/preputial abnormality - paraphimosis	1
Skeletal abnormality - congenital	1
Weight gain	1
Dyspnoea - with coughing	1
Mentation altered - depressed	1
Post-operative complication - other	1
Regurgitation	1
Tachypnoea	1

Table A6.2: Frequency of Syndromes With One Presenting Sign

Sign	Frequency
Faecal appearance abnormal - diarrhoea	1215
Vomiting - other	1094
Pruritus	931
Gait abnormality - lameness	792
Coughing	282
Ophthalmic (eye) abnormality	160
Ear (aural) abnormality	156

Continued on next page

Table A6.2: Frequency of Syndromes With One Presenting Sign (continued)

Sign	Frequency
Presenting complaint not listed	77
Wound	71
Skin (cutaneous) abnormality - other	57
Mass/swelling - skin (cutaneous)	56
Discharge - ocular (eye)	41
Skin (cutaneous) abnormality - eruptions/hives/rash	35
Anal irritation	34
Urination abnormal - other	33
Traumatic episode	31
Mass/swelling - other	28
Musculoskeletal injury	21
Dental (tooth) abnormality	18
Discharge - vaginal	16
Seizure(s)	16
Dietary indiscretion - other	13
Mass/swelling - eye (ophthalmic)	12
Urination abnormal - pollakiuria	12
Dietary indiscretion - foreign body ingestion	10
Eyelid abnormality	10
Red eye	10
Ophthalmic (eye) injury	9
Alopecia	8
Faecal appearance abnormal - haematochezia	8
Incontinence - urinary	8
Sneezing	8
Discharge - aural (ear)	7
Mass/swelling - abdominal	7
Vomiting - haematemesis	7
Appetite decreased	6
Behavioural abnormality	6
Mass/swelling - oral (mouth)	6

Continued on next page

Table A6.2: Frequency of Syndromes With One Presenting Sign (continued)

Sign	Frequency
Oral (mouth) abnormality - other	6
Pain - ophthalmic (eye)	6
Pyrexia/hyperthermia	6
Mass/swelling - ear (aural)	5
Pain - musculoskeletal	5
Post-operative complication - wound related	5
Abdominal distension	4
Collapse/syncopal episodes	4
Lethargy	4
Mass/swelling - musculoskeletal	4
Obesity	4
Salivation - increased/drooling	4
Oral (mouth) injury	3
Pain - on eating/chewing	3
Panting	3
Tremors/shaking/trembling	3
Cardiac (heart) abnormality - other	2
Faecal appearance abnormal - other	2
Intoxication (poisoning)/toxin exposure	2
Mentation altered - anxious/distressed	2
Pain - aural (ear)	2
Pain - oral (mouth)	2
Paresis/paralysis	2
Testicular/scrotal abnormality	2
Urination abnormal - dysuria	2
Urine colour abnormal	2
Weakness	2
Bleeding - epistaxis	1
Blindness	1
Cataract	1
Colic	1

Continued on next page

Table A6.2: Frequency of Syndromes With One Presenting Sign (continued)

Sign	Frequency
Defaecation abnormal - other	1
Discharge - nasal (nose)	1
Discharge - penile	1
Ear (aural) injury	1
Failure to gain weight	1
Gagging/retching	1
Gait abnormality - ataxia	1
Gait abnormality - other	1
Halitosis	1
Incontinence - faecal	1
Lymphadenomegaly	1
Mass/swelling - perianal	1
Pain - abdominal	1
Penile/preputial abnormality - paraphimosis	1
Polydipsia	1
Polyuria	1
Skeletal abnormality - congenital	1
Stiffness	1
Urination abnormal - inappropriate urination	1
Weight gain	1
Weight loss	1

Table A6.3: Frequency of Syndromes With Two Presenting Signs

Frequency	Sign 1	Sign 2
363	Faecal appearance abnormal - diarrhoea	Vomiting - other
25	Coughing	Vomiting - other
20	Pruritus	Vomiting - other
13	Dietary indiscretion - other	Faecal appearance abnormal - diarrhoea
13	Ear (aural) abnormality	Pruritus
13	Vomiting - haematemesis	Vomiting - other
9	Dietary indiscretion - other	Vomiting - other
6	Dietary indiscretion - foreign body ingestion	Vomiting - other
6	Presenting complaint not listed	Pruritus
5	Ear (aural) abnormality	Ophthalmic (eye) abnormality
5	Faecal appearance abnormal - diarrhoea	Faecal appearance abnormal - haematochezia
5	Pruritus	Skin (cutaneous) abnormality - other
4	Discharge - ocular (eye)	Ophthalmic (eye) abnormality
4	Gait abnormality - lameness	Traumatic episode
4	Mass/swelling - other	Pruritus
4	Ophthalmic (eye) abnormality	Ophthalmic (eye) injury
3	Discharge - nasal (nose)	Sneezing
3	Discharge - ocular (eye)	Ear (aural) abnormality
3	Faecal appearance abnormal - diarrhoea	Vomiting - haematemesis

Continued on next page

Table A6.3: Frequency of Syndromes With Two Presenting Signs (continued)

Frequency	Sign 1	Sign 2
3	Gait abnormality - lameness	Presenting complaint not listed
3	Gait abnormality - lameness	Pruritus
3	Gait abnormality - lameness	Wound
3	Mass/swelling - oral (mouth)	Vomiting - other
3	Polydipsia	Polyuria
3	Pruritus	Skin (cutaneous) abnormality - eruptions/hives/rash
3	Seizure(s)	Vomiting - other
2	Coughing	Presenting complaint not listed
2	Dietary indiscretion - other	Intoxication (poisoning)/toxin exposure
2	Discharge - ocular (eye)	Red eye
2	Ear (aural) abnormality	Pain - aural (ear)
2	Ear (aural) abnormality	Skin (cutaneous) abnormality - other
2	Faecal appearance abnormal - diarrhoea	Presenting complaint not listed
2	Gait abnormality - ataxia	Gait abnormality - lameness
2	Mass/swelling - skin (cutaneous)	Skin (cutaneous) abnormality - eruptions/hives/rash
2	Mentation altered - anxious/distressed	Presenting complaint not listed
2	Skin (cutaneous) abnormality - eruptions/hives/rash	Skin (cutaneous) abnormality - other

Continued on next page

Table A6.3: Frequency of Syndromes With Two Presenting Signs (continued)

Frequency	Sign 1	Sign 2
2	Skin (cutaneous) abnormality - other	Traumatic episode
2	Vomiting - other	Wound
1	Alopecia	Pruritus
1	Alopecia	Wound
1	Appetite decreased	Lethargy
1	Appetite decreased	Mentation altered - depressed
1	Appetite decreased	Pruritus
1	Appetite decreased	Salivation - increased/drooling
1	Appetite decreased	Weight loss
1	Coughing	Anal irritation
1	Coughing	Discharge - aural (ear)
1	Coughing	Dyspnoea - with coughing
1	Coughing	Gagging/retching
1	Coughing	Ophthalmic (eye) abnormality
1	Coughing	Sneezing
1	Coughing	Vomiting - haematemesis
1	Defaecation abnormal - defaecatory tenesmus	Defaecation abnormal - other
1	Defaecation abnormal - defaecatory tenesmus	Mass/swelling - perianal
1	Dental (tooth) abnormality	Pain - oral (mouth)
1	Dental (tooth) abnormality	Traumatic episode

Continued on next page

Table A6.3: Frequency of Syndromes With Two Presenting Signs (continued)

Frequency	Sign 1	Sign 2
1	Dietary indiscretion - foreign body ingestion	Faecal appearance abnormal - diarrhoea
1	Dietary indiscretion - foreign body ingestion	Oral (mouth) abnormality - other
1	Dietary indiscretion - foreign body ingestion	Oral (mouth) injury
1	Dietary indiscretion - other	Presenting complaint not listed
1	Dietary indiscretion - other	Vomiting - haematemesis
1	Discharge - aural (ear)	Ear (aural) injury
1	Discharge - ocular (eye)	Pruritus
1	Discharge - ocular (eye)	Skin (cutaneous) abnormality - eruptions/hives/rash
1	Discharge - ocular (eye)	Sneezing
1	Discharge - vaginal	Presenting complaint not listed
1	Discharge - vaginal	Pruritus
1	Discharge - vaginal	Urination abnormal - pollakiuria
1	Dry eye	Ophthalmic (eye) abnormality
1	Ear (aural) abnormality	Discharge - aural (ear)
1	Ear (aural) abnormality	Ear (aural) injury
1	Ear (aural) abnormality	Gait abnormality - lameness
1	Ear (aural) abnormality	Incontinence - urinary
1	Ear (aural) abnormality	Mass/swelling - ear (aural)

Continued on next page

Table A6.3: Frequency of Syndromes With Two Presenting Signs (continued)

Frequency	Sign 1	Sign 2
1	Ear (aural) abnormality	Skin (cutaneous) abnormality - eruptions/hives/rash
1	Ear (aural) abnormality	Tremors/shaking/trembling
1	Ear (aural) abnormality	Wound
1	Eyelid abnormality	Mass/swelling - eye (ophthalmic)
1	Eyelid abnormality	Mass/swelling - skin (cutaneous)
1	Faecal appearance abnormal - diarrhoea	Anal irritation
1	Faecal appearance abnormal - diarrhoea	Faecal appearance abnormal - other
1	Faecal appearance abnormal - diarrhoea	Gait abnormality - lameness
1	Faecal appearance abnormal - diarrhoea	Mass/swelling - other
1	Faecal appearance abnormal - diarrhoea	Salivation - increased/drooling
1	Faecal appearance abnormal - diarrhoea	Weight loss
1	Gait abnormality - lameness	Mass/swelling - musculoskeletal
1	Gait abnormality - lameness	Pain - musculoskeletal
1	Gait abnormality - lameness	Posture abnormal - other
1	Gait abnormality - lameness	Pyrexia/hyperthermia
1	Gait abnormality - lameness	Vomiting - other
1	Gait abnormality - other	Posture abnormal - other
1	Incontinence - urinary	Urination abnormal - inappropriate urination
1	Incontinence - urinary	Urination abnormal - pollakiuria

Continued on next page

Table A6.3: Frequency of Syndromes With Two Presenting Signs (continued)

Frequency	Sign 1	Sign 2
1	Lethargy	Vomiting - other
1	Lethargy	Weakness
1	Mass/swelling - abdominal	Mass/swelling - skin (cutaneous)
1	Mass/swelling - oral (mouth)	Mass/swelling - ear (aural)
1	Mass/swelling - other	Mass/swelling - skin (cutaneous)
1	Mass/swelling - other	Post-operative complication - wound related
1	Mass/swelling - other	Presenting complaint not listed
1	Mass/swelling - other	Traumatic episode
1	Mass/swelling - skin (cutaneous)	Presenting complaint not listed
1	Mass/swelling - skin (cutaneous)	Pruritus
1	Mass/swelling - skin (cutaneous)	Skin (cutaneous) abnormality - other
1	Ophthalmic (eye) abnormality	Pain - ophthalmic (eye)
1	Ophthalmic (eye) abnormality	Pruritus
1	Ophthalmic (eye) abnormality	Red eye
1	Ophthalmic (eye) injury	Mass/swelling - eye (ophthalmic)
1	Oral (mouth) abnormality - ulceration	Wound
1	Pain - other	Urination abnormal - pollakiuria
1	Penile/preputial abnormality - other	Skin (cutaneous) abnormality - eruptions/hives/rash
1	Polyuria	Urine colour abnormal

Continued on next page

Table A6.3: Frequency of Syndromes With Two Presenting Signs (continued)

Frequency	Sign 1	Sign 2
1	Post-operative complication - wound related	Pruritus
1	Presenting complaint not listed	Musculoskeletal injury
1	Presenting complaint not listed	Skin (cutaneous) abnormality - other
1	Presenting complaint not listed	Stiffness
1	Presenting complaint not listed	Urination abnormal - other
1	Presenting complaint not listed	Wound
1	Pruritus	Discharge - aural (ear)
1	Pruritus	Ear (aural) injury
1	Pruritus	Mass/swelling - ear (aural)
1	Pruritus	Mass/swelling - eye (ophthalmic)
1	Pruritus	Red eye
1	Pyrexia/hyperthermia	Vomiting - other
1	Red eye	Pain - ophthalmic (eye)
1	Skin (cutaneous) abnormality - eruptions/hives/rash	Vomiting - other
1	Skin (cutaneous) abnormality - other	Vomiting - other
1	Traumatic episode	Oral (mouth) injury
1	Traumatic episode	Wound
1	Urination abnormal - oliguria	Urination abnormal - pollakiuria
1	Urination abnormal - other	Urination abnormal - pollakiuria

Continued on next page

Table A6.3: Frequency of Syndromes With Two Presenting Signs (continued)

Frequency	Sign 1	Sign 2
1	Wound	Ear (aural) injury

Table A6.4: Frequency of Syndromes With Three Presenting Signs

Frequency	Sign 1	Sign 2	Sign 3
5	Faecal appearance abnormal - diarrhoea	Vomiting - haematemesis	Vomiting - other
3	Dietary indiscretion - other	Faecal appearance abnormal - diarrhoea	Vomiting - other
2	Coughing	Gagging/retching	Vomiting - other
2	Dietary indiscretion - foreign body ingestion	Faecal appearance abnormal - diarrhoea	Vomiting - other
1	Anorexia	Gait abnormality - lameness	Pyrexia/hyperthermia
1	Appetite decreased	Lethargy	Vomiting - other
1	Appetite decreased	Presenting complaint not listed	Anal irritation
1	Coughing	Dietary indiscretion - foreign body ingestion	Vomiting - other
1	Coughing	Ear (aural) abnormality	Pruritus
1	Coughing	Gagging/retching	Regurgitation
1	Coughing	Presenting complaint not listed	Vomiting - other
1	Coughing	Vomiting - haematemesis	Vomiting - other
1	Defaecation abnormal - other	Dietary indiscretion - other	Paresis/paralysis
1	Discharge - ocular (eye)	Ophthalmic (eye) abnormality	Pain - ophthalmic (eye)

Continued on next page

Table A6.4: Frequency of Syndromes With Three Presenting Signs (continued)

Frequency	Sign 1	Sign 2	Sign 3
1	Discharge - ocular (eye)	Ophthalmic (eye) abnormality	Red eye
1	Discharge - ocular (eye)	Red eye	Pain - ophthalmic (eye)
1	Ear (aural) abnormality	Ophthalmic (eye) abnormality	Ophthalmic (eye) injury
1	Ear (aural) abnormality	Pruritus	Discharge - aural (ear)
1	Faecal appearance abnormal - diarrhoea	Faecal appearance abnormal - haematochezia	Vomiting - haematemesis
1	Faecal appearance abnormal - diarrhoea	Faecal appearance abnormal - haematochezia	Vomiting - other
1	Faecal appearance abnormal - diarrhoea	Lymphadenomegaly	Vomiting - other
1	Faecal appearance abnormal - diarrhoea	Pyrexia/hyperthermia	Vomiting - other
1	Gait abnormality - lameness	Mass/swelling - skin (cutaneous)	Skin (cutaneous) abnormality - other
1	Gait abnormality - lameness	Pain - other	Post-operative complication - other
1	Gait abnormality - lameness	Pruritus	Vomiting - other
1	Gait abnormality - lameness	Pyrexia/hyperthermia	Tachypnoea

Continued on next page

Table A6.4: Frequency of Syndromes With Three Presenting Signs (continued)

Frequency	Sign 1	Sign 2	Sign 3
1	Mass/swelling - other	Pruritus	Skin (cutaneous) abnormality - other
1	Mass/swelling - other	Pruritus	Vomiting - other
1	Presenting complaint not listed	Pruritus	Vomiting - other
1	Urination abnormal - dysuria	Urination abnormal - pollakiuria	Urine colour abnormal
1	Urination abnormal - oliguria	Urination abnormal - pollakiuria	Urine colour abnormal

One syndrome had four signs and one had five. They were each reported just once.

Four Signs Coughing; Discharge - ocular (eye); Ophthalmic (eye) abnormality; Vomiting - other

Five Signs Discharge - vaginal; Incontinence - urinary; Polydipsia; Polyuria; Urination abnormal - other

Appendix 7: Accelerometer Questionnaire

Dogslife ROSLIN THE UNIVERSITY of EDINBURGH The Royal (Dick) School of Veterinary Studies

Movement Recording

Thank you for helping the Dogslife project.

We are looking at the movement of dogs in the Dogslife project and how that relates to their daily activity. The information we collect will allow us to assess how different activities affect the health of dogs in their lifetime.

We have **enclosed a movement monitor** (accelerometer) for you to attach to your dog's collar. **This should be attached near to the buckle or clasp of your dog's collar** using the Velcro strap, as shown in the photographs below. The monitor is extremely small and light and your dog should not notice that it is there. The device records information every time your dog moves and the type of movement they do (don't worry if your dog goes into the water or gets it wet; the device is waterproof). We would be grateful if you could **leave the monitor on your dog's collar for a full seven days** before returning it to us in the pre-paid envelope provided. We would also be grateful if you could **complete and return the enclosed questionnaire** at the same time as you return the monitor.

We plan to discuss/publish information obtained in the study only in the form of grouped data. Any results from the study will be made available on the Dogslife website. We will never identify individual dogs in publications. If you have any questions about the activity monitor, please do not hesitate to contact us at info@dogslife.ac.uk, or telephone us on 0131 651 7309.

Thank you again for your time and help with this project, it is greatly appreciated.

Best wishes,

The Dogslife Team

c/o Erica Rose
Hospital for Small Animals
Royal (Dick) School of Veterinary Studies
and Roslin Institute,
The University of Edinburgh,
Easter Bush Veterinary Centre,
Roslin,
Midlothian,
EH25 9RG.

Attach the monitor to the collar using the Velcro strap provided.

The monitor should be attached to the outside surface of the collar near the buckle or clasp so that it sits at the very bottom of your dog's neck. Please be sure to wrap the Velcro strap tightly around the collar onto itself to secure it.

DIRECTOR OF THE ROSLIN INSTITUTE Professor David Hume FRSE FSB, Professor of Mammalian Functional Genetics, RDSVS
The Roslin Institute, a world class research centre, is incorporated with the Royal (Dick) School of Veterinary Studies, the University of Edinburgh
The Roslin Institute receives strategic funding from the BBSRC
The University of Edinburgh is a charitable body, registered in Scotland, with registration number SC005336

BBSRC bioscience for the future

Figure A12: Accelerometer cover letter with instructions for attaching the device

Dogslife

Using the movement monitor

- Attach the monitor near to the buckle or clasp of your dog's collar using the Velcro strap provided
- Make a note of the exact time and date you attach the monitor to your dog's collar and email the details to info@dogslife.ac.uk
- Leave the monitor on your dog's collar for a full seven days
- Remove the monitor from your dog's collar (and discard the Velcro strap) and return the monitor with the completed questionnaire in the pre-paid envelope provided
- Make a note of the exact time and date you removed the monitor from your dog's collar and email the details to info@dogslife.ac.uk

If you have any queries, please contact us at **info@dogslife.ac.uk**.
Thank you for helping Dogslife!

Figure A13: Accelerometer A5 flyer with instructions on the study protocol

Dogslife ROSLIN THE UNIVERSITY of EDINBURGH The Royal (Dick) School of Veterinary Studies

Questionnaire

On average, **in the last week (when your dog has been wearing the movement monitor), how long** has your dog spent doing the following exercise(s) **EACH DAY** (please tick)?

	on Weekdays	**on Weekend days**
Walking on the lead:	☐ None	☐ None
	☐ 1 - 5 minutes	☐ 1 - 5 minutes
	☐ 5 - 15 minutes	☐ 5 - 15 minutes
	☐ 15 - 30 minutes	☐ 15 - 30 minutes
	☐ 30 - 60 minutes	☐ 30 - 60 minutes
	☐ 1 - 2 hours	☐ 1 - 2 hours
	☐ over 2 hours	☐ over 2 hours
Running on the lead:	☐ None	☐ None
	☐ 1 - 5 minutes	☐ 1 - 5 minutes
	☐ 5 - 15 minutes	☐ 5 - 15 minutes
	☐ 15 - 30 minutes	☐ 15 - 30 minutes
	☐ 30 - 60 minutes	☐ 30 - 60 minutes
	☐ 1 - 2 hours	☐ 1 - 2 hours
	☐ over 2 hours	☐ over 2 hours
Walking/running off the lead:	☐ None	☐ None
	☐ 1 - 5 minutes	☐ 1 - 5 minutes
	☐ 5 - 15 minutes	☐ 5 - 15 minutes
	☐ 15 - 30 minutes	☐ 15 - 30 minutes
	☐ 30 - 60 minutes	☐ 30 - 60 minutes
	☐ 1 - 2 hours	☐ 1 - 2 hours
	☐ over 2 hours	☐ over 2 hours
Exercise involving fetching, chasing or retrieving:	☐ None	☐ None
	☐ 1 - 5 minutes	☐ 1 - 5 minutes
	☐ 5 - 15 minutes	☐ 5 - 15 minutes
	☐ 15 - 30 minutes	☐ 15 - 30 minutes
	☐ 30 - 60 minutes	☐ 30 - 60 minutes
	☐ 1 - 2 hours	☐ 1 - 2 hours
	☐ over 2 hours	☐ over 2 hours

continued overleaf....

Figure A14: First page of questionnaire sent with accelerometer

	Weekdays	Weekend days
Obedience Training:	☐ None	☐ None
	☐ 1 - 5 minutes	☐ 1 - 5 minutes
	☐ 5 - 15 minutes	☐ 5 - 15 minutes
	☐ 15 - 30 minutes	☐ 15 - 30 minutes
	☐ 30 - 60 minutes	☐ 30 - 60 minutes
	☐ 1 - 2 hours	☐ 1 - 2 hours
	☐ over 2 hours	☐ over 2 hours
Other playing activity (including dogs playing together)	☐ None	☐ None
	☐ 1 - 5 minutes	☐ 1 - 5 minutes
	☐ 5 - 15 minutes	☐ 5 - 15 minutes
	☐ 15 - 30 minutes	☐ 15 - 30 minutes
	☐ 30 - 60 minutes	☐ 30 - 60 minutes
	☐ 1 - 2 hours	☐ 1 - 2 hours
	☐ over 2 hours	☐ over 2 hours

In total, how much time has your dog spent travelling in a moving vehicle since the monitor was fitted?

Does this questionnaire encompass all the activity that your dog does each day (if not please detail the type and length of any additional activity (excluding sleeping and lying around) that your dog does each day)?

Attach id sticker

Thank you for completing the questionnaire; **please return this sheet with the monitor** (please **remove and discard the Velcro strip** and send us back the monitor on its own) in the pre-paid stamped addressed envelope provided.

If you have any queries, please do not hesitate to contact us by e-mail info@dogslife.ac.uk, or telephone us on 0131 651 7309. Thank you again for your time and help with this project.

Figure A15: Second page of questionnaire sent with accelerometer. Prior to posting, the project administrator would place a sticky label with the dog's ID in the lower box.

PHILIPS

Powerful, fast, and flexible physical activity monitoring

PHILIPS RESPIRONICS

Specifications

Size	1.14 in x 1.45 in x 0.43 without standard band 29 mm x 37 mm x 11 mm	
Weight	0.56 ounces, 16 grams, without standard band 0.77 ounces, 22 grams, with standard band	
Case materials	Polyurethane/polyester alloy	
Frame and battery cover	Titanium	
Attachment options	Wrist: Nylon wrist band (standard); elastic band (optional) Waist: Belt clip or waist band (optional) Ankle: Elastic band (optional)	
Battery	CR2025 lithium coin cell (user replaceable)	
Memory capacity	32 MB	
Maximum continuous recording time	Raw mode 12 days	Epoch mode 1 second + steps 194 days

Accelerometer details

Range	0.05 G to 2 G
Bandwidth	.035 Hz to 3.5 Hz
Resolution	100 counts or 0.02 G (at 1 G peak)
Sampling rate	32 Hz
Sampling modes	Raw + steps
	Epoch: 1, 2, 5, 15, 30, 60 sec Epoch + steps: 1, 2, 5, 15, 30, 60 sec

Environmental attributes

Moisture protection	Waterproof IEC60529 IPX7 1 meter for 30 min
Storage/transportation temperature range	-4 to 140°F (-20 to 60°C) to 95% humidity
Operating temperature range	41 to 104°F (5 to 40°C) 15% to 95% humidity

Computer attributes

Communication interface	9-pin RS-232 serial port (standard with ActiReader) 9-pin serial to USB adapter (optional)
Operating system compatibility	Windows XP, WindowsVista (32 and 64 bit), Windows 7 (32 and 64 bit)
Hardware platform	Personal computer

All products are not available in all countries. Please contact your sales representative or our customer service team for more information.

http://www.healthcare.philips.com/pwc_hc/main/homehealth/sleep/actical/

Figure A16: Specifications of the Philips Respironics Acti*Cal*

Appendix 8: Limber Tail Questionnaires

ROSLIN

Limber Tail Questionnaire

Thank you for participating in the Dogslife project. We are currently investigating factors that cause limber tail (and also known as cold tail, swimmers tail, frozen tail, rudder tail or limp tail) in Labrador Retrievers. We would like to find out more about the condition. As your dog is reported to have had this condition we would be very grateful if you could detail the following information about your dog, and return the questionnaire to us by e-mail (info@dogslife.ac.uk) or in the postal envelope provided. Thank you for your time and help.

(* = please circle the most appropriate answer)

1. **How many separate episodes of limber tail** has your dog had? …….

2. **Do the episodes follow?***

   a. Swimming? *Yes / No*
   b. Cold Weather? *Yes / No*
   c. Wet weather? *Yes / No*
   d. Vigorous exercise? *Yes / No*
   e. Resting in a confined area (e.g. dog crate) *Yes / No*
   f. Anything else? ……………………………………………………

3. **What does your dog's tail look like when the episodes occur?***

   a. Limp at the end *Yes / No*
   b. Limp along the entire length *Yes / No*
   c. Stiff at the base (near the body) *Yes / No*
   d. The hair on the top of it stands on end *Yes / No*
   e. It appears painful for no reason *Yes / No*

4. On average, **how long does the episode last?*** *An hour / a few hours / a day / a few days / a week or more*

5. **Can you avoid the episodes occurring**?* *Yes / No*

   If yes, how? ……………………………………………………………………………………………

6. On a scale of 0 to 10, **how painful would you say each episode is?**

   (0 = not painful, 10 = could not be more painful) …….

7. On a scale of 0 to 10, **how much does this condition affect your dog's quality of life?**

   (0 = does not affect my dogs quality of life, 10 = my dog's quality of life could not be worse, because of this condition)? …….

8. **Is there anything else you would like to tell us about this condition**?

   ……………………………………………………………………………………………………

9. **Does your dog go swimming?*** *Yes / No*

   If YES, **when** (all the year or just in summer), **how often** (daily or occasionally) and **where** (sea, canals, rivers etc.) does your dog go swimming?

   ……………………………………………………………………………………………………

Thank you for this extra information about your dog; this is a tremendous help to the Dogslife project.

The Roslin Institute, a world class research centre, is incorporated with the Royal (Dick) School of Veterinary Studies, the University of Edinburgh
The Roslin Institute receives strategic funding from the BBSRC
The University of Edinburgh is a charitable body, registered in Scotland, with registration number SC005336

Figure A17: Questionnaire sent to owners of suspected limber tail cases

ROSLIN

THE UNIVERSITY OF EDINBURGH

Tail Health Questionnaire

Thank you for participating in the Dogslife project. We are currently investigating factors that cause limber tail (and also known as cold tail, swimmers tail, frozen tail, rudder tail or limp tail) in Labrador Retrievers - to find out more about the condition you can read our previous newsletter article about it here: http://www.dogslife.ac.uk/newsletter/view/17#tail. As your dog has never been reported to Dogslife as having had this condition, we wanted to double-check some information with you so that we can compare your dog to others who are affected. We would be very grateful if you could complete the questionnaire and return it to us by e-mail (info@dogslife.ac.uk) or in the postal envelope provided. Thank you for your time and help.

(* = please circle the most appropriate answer)

1. **Does your dog go swimming?*** *Yes / No*

   If YES, **when** (e.g. all the year or just in summer), **how often** (e.g. daily or occasionally) and **where** (e.g. sea, canals, rivers etc.) does your dog go swimming?

   ………

2. **Has your dog ever had any episodes where they display any of the following signs with their tail?***

   a. It looks abnormally limp at the end *Yes / No*
   b. It looks abnormally limp along the entire length *Yes / No*
   c. It looks abnormally stiff at the base (near the body) *Yes / No*
   d. The hair on the top of it stands on end *Yes / No*
   e. It appears painful for no reason *Yes / No*

If you answered NO to ALL of the responses in Question 2, you have completed the questionnaire
If you answered YES to ANY of the responses in Question 2, please answer all the following questions;

3. On average, **how long does an episode last?*** *An hour / a few hours / a day / a few days / a week*

4. **Do the episodes follow?***

   a. Swimming? *Yes / No*
   b. Cold Weather? *Yes / No*
   c. Wet weather? *Yes / No*
   d. Vigorous exercise? *Yes / No*
   e. Resting in a confined area (e.g. dog crate) *Yes / No*
   f. Anything else? ………………………………………………

5. **Can you avoid the episodes occurring?*** *Yes / No*

   If yes, how? ………………………………………………………………………………………

6. On a scale of 0 to 10, **how painful would you say each episode is?**

   (0 = not painful, 10 = could not be more painful) …….

7. On a scale of 0 to 10, **how much does this condition affect your dog's quality of life?**

   (0 = does not affect my dogs quality of life, 10 = my dog's quality of life could not be worse, because of this condition)? …….

8. **Is there anything else you would like to tell us about this condition**?

   ………

Thank you for this extra information about your dog; this is a tremendous help to the Dogslife project.

The Roslin Institute, a world class research centre, is incorporated with the Royal (Dick) School of Veterinary Studies, the University of Edinburgh
The Roslin Institute receives strategic funding from the BBSRC
The University of Edinburgh is a charitable body, registered in Scotland, with registration number SC005336

BBSRC

Figure A18: Questionnaire sent to dogs with no report of limber tail

References

D. Alderton and T. Morgan. *Dogs*. Dorling Kindersley, London, 1993. ISBN 1564581799. 122

A. Alonso, M. Seguí-Gómez, J. de Irala, A. Sánchez-Villegas, J. J. Beunza, and M. A. Martínez-Gonzalez. Predictors of follow-up and assessment of selection bias from dropouts using inverse probability weighting in a cohort of university graduates. *European journal of epidemiology*, 21(5):351–358, Jan. 2006. ISSN 0393-2990. doi: 10.1007/s10654-006-9008-y. URL http://www.ncbi.nlm.nih.gov/pubmed/16736275. 36

ALSPAC. ALSPAC Bio Resource, 2015. URL http://www.bristol.ac.uk/alspac/researchers/resources-available/data-details/bio-resource/. 15

American Kennel Club. American Kennel Club Labrador Retriever Breed Standard, 2014. URL https://www.akc.org/breeds/labrador_retriever/breed_standard.cfmhttp://www.webcitation.org/6UUdCtIZX. 140

E. Banks, V. Beral, G. Reeves, A. Balkwill, and I. Barnes. Fracture incidence in relation to the pattern of use of hormone therapy in postmenopausal women. *The Journal of the American Medical Association*, 291(18):2212–2220, 2004. URL http://jama.ama-assn.org/content/291/18/2212.short. 5

P. C. Bartlett, J. W. Van Buren, M. Neterer, and C. Zhou. Disease surveillance and referral bias in the veterinary medical database. *Preventive veterinary medicine*, 94(3-4):264–71, May 2010. ISSN 1873-1716. doi: 10.1016/j.prevetmed.2010.01.007. URL http://www.ncbi.nlm.nih.gov/pubmed/20129684. 7

K. Bartoń. MuMIn: Multi-model inference, 2014. URL http://cran.r-project.org/package=MuMIn. 90, 193

D. M. Bates, M. Maechler, B. M. Bolker, and S. Walker. lme4: Linear mixed-models using Eigen and S4., 2014. URL http://cran.r-project.org/package=lme4. 83, 193

R. Bivand and N. Lewin-Koh. maptools: Tools for Reading and Handling Spatial Objects, 2015. URL http://cran.r-project.org/package=maptools. 148

B. N. Bonnett and A. Egenvall. Age patterns of disease and death in insured Swedish dogs, cats and horses. *Journal of Comparative Pathology*, 142:S33–S38, Jan. 2010. ISSN 1532-3129. doi: 10.1016/j.jcpa.2009.10.008. URL http://www.ncbi.nlm.nih.gov/pubmed/19932895. 7

B. N. Bonnett, A. Egenvall, P. Olson, and A. Hedhammar. Mortality in insured Swedish dogs: rates and causes of death in various breeds. *Veterinary Record*, 141(2):40–44, July 1997. ISSN 0042-4900. doi: 10.1136/vr.141.2.40. URL http://veterinaryrecord.bmj.com/cgi/doi/10.1136/vr.141.2.40. 14

I. Bray, S. Noble, A. Boyd, L. Brown, P. Hayes, J. Malcolm, R. Robinson, R. Williams, K. Burston, J. Macleod, L. Molloy, and K. Tilling. A randomised controlled trial comparing opt-in and opt-out home visits for tracing lost participants in a prospective birth cohort study. *BMC Medical Research Methodology*, 15(1):52, 2015. ISSN 1471-2288. doi: 10.1186/s12874-015-0041-y. URL http://www.biomedcentral.com/1471-2288/15/52. 244

W. Bredal and S. Thoresen. “Water tail”, a little known condition in dogs. *Norsk Veterinaertidsskrift*, 111(7):479–480, 1999. 217

E. A. Bruno, J. W. Guthrie, S. A. Ellwood, R. J. Mellanby, and D. N. Clements. Global Positioning System Derived Performance Measures Are Responsive Indicators of Physical Activity, Disease and the Success of Clinical Treatments in Domestic Dogs. *Plos One*, 10(2):e0117094, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0117094. URL http://dx.plos.org/10.1371/journal.pone.0117094. 187

D. N. Clements, I. G. Handel, E. Rose, D. Querry, C. A. Pugh, W. E. Ollier, K. L. Morgan, L. J. Kennedy, J. Sampson, K. M. Summers, and B. M. C. de Bronsvoort. Dogslife: A web-based longitudinal study of Labrador Retriever health in the UK. *BMC veterinary research*, 9:13, Jan. 2013. ISSN 1746-6148. doi: 10.1186/1746-6148-9-13. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3559277&tool=pmcentrez&rendertype=abstract. 13, 17, 19, 20, 25, 36, 76, 78, 80, 81

O. Courtenay, D. W. Macdonald, R. Lainson, J. J. Shaw, and C. Dye. Epidemiology of canine leishmaniasis: a comparative serological study of dogs and foxes in Amazon Brazil. *Parasitology*, 109(3):273–279, Sept. 1994. ISSN 0031-1820. URL http://www.ncbi.nlm.nih.gov/pubmed/7970884. 9

W. A. Coward. The doubly-labelled-water (2H218O) method: principles and practice. *Proceedings of the Nutrition Society*, 47(3):209–218, 1988. 187

D. R. Cox. Regression Models and Life Tables (with Discussion). *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972. 24, 147

F. Cunningham, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Giron, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, A. K. Kahari, S. Keenan, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, B. Overduin, A. Parker, M. Patricio, E. Perry, M. Pignatelli, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, B. L. Aken, E. Birney, J. Harrow, R. Kinsella, M. Muffato, M. Ruffier, S. M. J. Searle, G. Spudich, S. J. Trevanion, A. Yates, D. R. Zerbino, and P. Flicek. Ensembl 2015. *Nucleic Acids Research*, 43 (D1):D662–D669, 2014. ISSN 0305-1048. doi: 10.1093/nar/gku1010. URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku1010. 222

T. Curtin, S. Ingels, S. Wu, R. Heuer, and J. Owings. National longitudinal study of 1988: Base year to fourth follow-up data file users manual (NCES 2002-323). Technical report, US Department of Education, National Center for Education Statistics, Washington DC, 2002. 242

A. Custovic and A. Simpson. The Role of Inhalant Allergens in Allergic Airways Disease. *Journal of Investigational Allergology and Clinical Immunology*, 22(2):

393–401, 2012. URL http://www.jiaci.org/issues/vol22issue6/1.pdf. 139

G. Dahlquist, W. Stig, J.-I. Ivarsson, G. Sterky, and K. Tengvald. Health Problems and Care in Young Families - An Evaluation of Survey Procedures. *International Journal of Epidemiology*, 13(2):221–228, 1984. 180

C. Dancey and J. Reidy. *Statistics without maths for psychology: using SPSS for windows*. Prentice Hall, London, 2004. 192, 206, 208, 210

J. L. Daniels, M. P. Longnecker, A. S. Rowland, and J. Golding. Fish Intake During Pregnancy and Early Cognitive Development of Offspring. *Epidemiology*, 15(4):394–402, July 2004. ISSN 1044-3983. doi: 10.1097/01.ede.0000129514.46451.ce. URL http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00001648-200407000-00004. 5

I. J. Deary, A. J. Gow, A. Pattie, and J. M. Starr. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *International Journal of Epidemiology*, 41 (6):1576–1584, Dec. 2012. ISSN 1464-3685. doi: 10.1093/ije/dyr197. URL http://www.ncbi.nlm.nih.gov/pubmed/22253310. 5

J. Dobson, T. Hoather, T. J. McKinley, and J. L. N. Wood. Mortality in a cohort of flat-coated retrievers in the UK. *Veterinary and Comparative Oncology*, 7(2): 115–121, Feb. 2009. ISSN 14765810. doi: 10.1111/j.1476-5829.2008.00181.x. URL http://doi.wiley.com/10.1111/j.1476-5829.2008.00181.x. 10

Dogslife. Dogslife, 2015. URL http://www.dogslife.ac.uk. 1, 13

I. Dohoo, W. Martin, and H. Stryhn. Introduction to observational studies. In *Veterinary Epidemiologic Research*, chapter 7, pages 151–166. VER Inc, Charlottetown, 2nd edition, 2010a. ISBN 978-0-919013-60-5. 1

I. Dohoo, W. Martin, and H. Stryhn. Cohort Studies. In *Veterinary Epidemiologic Research*, chapter 8, pages 167–180. VER Inc, Charlottetown, 2 edition, 2010b. ISBN 978-0-919013-60-5. 6, 9

I. Dohoo, W. Martin, and H. Stryhn. Controlled Studies. In I. Dohoo, editor, *Veterinary Epidemiologic Research*, chapter 11, pages 213–235. VER Inc, Charlottetown, 2nd edition, 2010c. ISBN 978-0-919013-60-5. 24

R. Doll. Tobacco: a medical history. *Journal of Urban Health*, 76(3):289–313, 1999. URL http://link.springer.com/article/10.1007/BF02345669. 4

R. Doll, R. Peto, J. Boreham, and I. Sutherland. Mortality in relation to smoking: 50 years' observations on male British doctors. *British Medical Journal*, 328(7455):1519, June 2004. ISSN 1756-1833. doi: 10.1136/bmj.38142.554479.AE. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=437139&tool=pmcentrez&rendertype=abstract. 4

C. Dow, K. E. Michel, M. Love, and D. C. Brown. Evaluation of optimal sampling interval for activity monitoring in companion dogs. *American Journal of Veterinary Research*, 70(4):444–448, 2009. 187, 189, 213

A. T. Edney and P. M. Smith. Study of obesity in dogs visiting veterinary practices in the United Kingdom. *Veterinary Record*, 118(14):391–396, 1986. URL http://veterinaryrecord.bmj.com/content/118/14/391.abstract. 140

P. J. Edwards, I. G. Roberts, M. J. Clarke, C. DiGuiseppi, R. Wentz, I. Kwan, R. Cooper, L. Felix, and S. Pratap. Methods to increase response rates to postal and electronic questionnaires (Review). *Cochrane Database Systematic Reviews*, 3(MR000008), 2009. 36

A. Egenvall, B. N. Bonnett, P. Olson, and A. Hedhammar. Validation of computerized Swedish dog and cat insurance data against veterinary practice records. *Preventive Veterinary Medicine*, 36(1):51–65, July 1998. ISSN 0167-5877. URL http://www.ncbi.nlm.nih.gov/pubmed/9677627. 7

A. Egenvall, B. N. Bonnett, M. Shoukri, P. Olson, A. Hedhammar, and I. Dohoo. Age pattern of mortality in eight breeds of insured dogs in Sweden. *Preventive veterinary medicine*, 46(1):1–14, July 2000. ISSN 0167-5877. URL http://www.ncbi.nlm.nih.gov/pubmed/10854932. 179

A. Egenvall, B. N. Bonnett, P. Ohagen, P. Olson, A. Hedhammar, and H. von Euler. Incidence of and survival after mammary tumors in a population of over 80,000 insured female dogs in Sweden from 1995 to 2002. *Preventive Veterinary Medicine*, 69(1-2):109–127, June 2005. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2005.01.014. URL http://www.ncbi.nlm.nih.gov/pubmed/15899300. 7

S. Elnegaard, R. S. Andersen, A. F. Pedersen, P. V. Larsen, J. Søndergaard, S. Rasmussen, K. Balasubramaniam, R. P. Svendsen, P. Vedsted, and D. E. Jarbøl. Self-reported symptoms and healthcare seeking in the general population -exploring "The Symptom Iceberg". *BMC Public Health*, 15:685, 2015. ISSN 1471-2458. doi: 10.1186/s12889-015-2034-5. URL http://www.biomedcentral.com/1471-2458/15/685. 179, 180, 183

L. L. Farrell, J. J. Schoenebeck, P. Wiener, D. N. Clements, and K. M. Summers. The challenges of pedigree dog health: approaches to combating inherited disease. *Canine Genetics and Epidemiology*, 2(3):1–14, 2015. ISSN 2052-6687. doi: 10.1186/s40575-015-0014-9. URL http://www.cgejournal.org/content/2/1/3. 253

M. P. Fay. Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics*, 11:373–374, 2010. 220

D. R. Feikin, A. Audi, B. Olack, G. M. Bigogo, C. Polyak, H. Burke, J. Williamson, and R. F. Breiman. Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. *International journal of epidemiology*, 39(2):450–458, Apr. 2010. ISSN 1464-3685. doi: 10.1093/ije/dyp374. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2846445&tool=pmcentrez&rendertype=abstract. 81

R. A. Fisher. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society*, 85(1):87–94, 1922. 24

A. Fraser, C. Macdonald-Wallis, K. Tilling, A. Boyd, J. Golding, G. Davey smith, J. Henderson, J. Macleod, L. Molloy, A. Ness, S. Ring, S. M. Nelson, and D. A. Lawlor. Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. *International Journal of Epidemiology*, 42(1):97–110, 2013. ISSN 03005771. doi: 10.1093/ije/dys066. 4

M. Frischknecht, H. Niehof-Oellers, V. Jagannathan, M. Owczarek-Lipska, C. Drögemüller, E. Dietschi, G. Dolf, B. Tellhelm, J. Lang, K. Tiira, H. Lohi, and T. Leeb. A COL11A2 mutation in Labrador retrievers with mild disproportionate dwarfism. *PloS one*, 8(3):

e60149, Jan. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0060149. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3603880&tool=pmcentrez&rendertype=abstract. 141

L. T. Glickman, N. W. Glickman, G. E. Moore, E. M. Lund, G. C. Lantz, and B. M. Pressler. Association between chronic azotemic kidney disease and the severity of periodontal disease in dogs. *Preventive Veterinary Medicine*, 99(2-4):193–200, May 2011. ISSN 1873-1716. doi: 10.1016/j.prevetmed.2011.01.011. URL http://www.ncbi.nlm.nih.gov/pubmed/21345505. 7

J. Golding. Children of the nineties. A longitudinal study of pregnancy and childhood based on the population of Avon. *West of England Medical Journal*, 105(3):80–2, 1990. 4, 242

J. Golding, P. Fleming, and S. Parkes. Cot deaths and sleep position campaigns. *The Lancet*, 339(8795):748–749, 1992. URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Cot+deaths+and+sleep+position+campaigns.#0. 5

M. Gramiccia, T. Di Muccio, E. Fiorentino, A. Scalone, G. Bongiorno, S. Cappiello, R. Paparcone, V. Foglia Manzillo, M. Maroli, L. Gradoni, and G. Oliva. Longitudinal study on the detection of canine Leishmania infections by conjunctival swab analysis and correlation with entomological parameters. *Veterinary Parasitology*, 171(3-4):223–228, Aug. 2010. ISSN 1873-2550. doi: 10.1016/j.vetpar.2010.03.025. URL http://www.ncbi.nlm.nih.gov/pubmed/20409639. 9

G. C. Gray, K. B. Chesbrough, M. A. Ryan, P. Amoroso, E. J. Boyko, G. D. Gackstetter, T. I. Hooper, and J. R. Riddle. The millennium Cohort Study: a 21-year prospective cohort study of 140,000 military personnel. *Military Medicine*, 167(4):483–488, 2002. URL http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA412574. 6

E. H. Greeley, E. Spitznagel, D. F. Lawler, R. D. Kealy, and M. Segre. Modulation of canine immunosenescence by life-long caloric restriction. *Veterinary Immunology and Immunopathology*, 111(3-4):287–299, June 2006. ISSN 0165-2427. doi: 10.1016/j.vetimm.2006.02.002. URL http://www.ncbi.nlm.nih.gov/pubmed/16567002. 9

M. K. Guy, R. L. Page, W. A. Jensen, P. N. Olson, J. D. Haworth, E. E. Searfoss, and D. E. Brown. The Golden Retriever Lifetime Study: establishing an observational cohort study with translational relevance for human health. *Philosophical Transactions of the Royal Society B*, 370, 2015. doi: 10.1098/rstb.2014.0230. URL http://rstb.royalsocietypublishing.org/content/370/1673/20140230.long#sec-10. 13, 245, 253

B. D. Hansen, B. D. X. Lascelles, B. W. Keene, A. K. Adams, and A. E. Thomson. Monitoring of Spontaneous Activity in Dogs. *American Journal of Veterinary Research*, 68(5):468–475, 2007. ISSN 00029645. doi: 10.2460/ajvr.68.5.468. 190, 213

C. A. Hercock, G. L. Pinchbeck, A. Giejda, P. D. Clegg, and J. F. Innes. Validation of a client-based clinical metrology instrument for the evaluation of canine elbow osteoarthritis. *The Journal of small animal practice*, 50(6):266–71, June 2009. ISSN 1748-5827. doi: 10.1111/j.1748-5827.2009.00765.x. URL http://www.ncbi.nlm.nih.gov/pubmed/19527419. 37

C. Hewison. Frozen tail or limber tail in working dogs. *Veterinary Record*, 140: 536, 1997. 217

J. V. Hinrichs. A two-process memory-strength theory for judgement of recency. *Psychological Review*, 77(3):223–233, 1970. URL http://psycnet.apa.org/?fa=main.doiLanding&uid=1970-11793-001. 80

J. W. Hotchkiss, S. W. Reid, and R. Christley. Construction and validation of a risk-screening questionnaire for the investigation of recurrent airway obstruction in epidemiological studies of horse populations in Great Britain. *Preventive veterinary medicine*, 75(1-2):8–21, July 2006. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2006.01.001. URL http://www.ncbi.nlm.nih.gov/pubmed/16488493. 37

J. E. F. Houlton. A survey of gundog lameness and injuries in Great Britain in the shooting seasons 2005/2006 and 2006/2007. *Veterinary and Comparative Orthopaedics and Traumatology*, 21(3):231–237, 2008. ISSN 09320814. 218

K. Hubbard, B. J. Skelly, J. McKelvie, and J. L. N. Wood. Risk of vomiting and diarrhoea in dogs. *Veterinary*, 161(22):755–757, 2007. URL http://veterinaryrecord.bmj.com/content/161/22/755. 155, 182, 183, 184

J. L. Huck, D. N. Biery, D. F. Lawler, T. P. Gregor, J. J. Runge, R. H. Evans, R. D. Kealy, and G. K. Smith. A longitudinal study of the influence of lifetime food restriction on development of osteoarthritis in the canine elbow. *Veterinary Surgery*, 38(2):192–198, Feb. 2009. ISSN 1532-950X. doi: 10.1111/j.1532-950X.2008.00487.x. URL http://www.ncbi.nlm.nih.gov/pubmed/19236677. 10

L. Hunt, P. Fleming, and J. Golding. Does the Supine Sleeping Position Have Any Adverse Effects on the Child?: I. Health in the First Six Months. *Pediatrics*, 100(1):e11–e11, July 1997. ISSN 0031-4005. doi: 10.1542/peds.100.1.e11. URL http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.100.1.e11. 5

A. Indrebø, C. Trangerud, and L. Moe. Canine neonatal mortality in four large breeds. *Acta Veterinaria Scandinavica*, 49(Suppl 1):S2, 2007. ISSN 1751-0147. doi: 10.1186/1751-0147-49-S1-S2. URL http://www.actavetscand.com/content/49/S1/S2. 13, 243

J. F. Innes and A. R. Barr. Can owners assess outcome following treatment of canine cruciate ligament deficiency? *The Journal of small animal practice*, 39 (8):373–378, 1998. ISSN 0022-4510. doi: 10.1111/j.1748-5827.1998.tb03735.x. 187

D. James and S. DebRoy. RMySQL: R interface to the MySQL database, 2012. URL http://cran.r-project.org/package=RMySQL. 24

W. Jeffels. Frozen tail or limber tail in working dogs. *Veterinary Record*, 140: 564, 1997. 217

M. Jeffreys, K. Northstone, J. Holly, P. Emmett, and D. Gunnell. Levels of insulin-like growth factor during pregnancy and maternal cancer: a nested case-control study. *Cancer Causes & Control*, 22(7):945–953, 2011. 14

R. Kaaks and E. Riboli. Validation and Calibration of Dietary Intake Measurements in the EPIC Project : Methodological Considerations. *International Journal of Epidemiology*, 26(1):15–25, 1997. 37

R. Kealy, D. Lawler, J. Ballam, G. Lust, G. Smith, D. Biery, and S. Olsson. Five-year longitudinal study on limited food consumption and development of osteoarthritis in coxofemoral joints of dogs. *Journal of the American Veterinary Medical Association*, 210(2):222–225, 1997. 10

R. D. Kealy, S. Olsson, K. Monti, D. F. Lawler, D. N. Biery, R. Helms, G. Lust, and G. K. Smith. Effects of limited food consumption on the incidence of hip dysplasia in growing dogs. *Journal of the American Veterinary Medical Association*, 201(6):857–863, 1992. 9

R. D. Kealy, D. F. Lawler, J. M. Ballam, G. Lust, D. N. Biery, G. K. Smith, and S. L. Mantz. Evaluation of the effect of limited food consumption on radiographic evidence of osteoarthritis in dogs. *Journal of the American Veterinary Medical Association*, 217(11):1678–1680, Dec. 2000. ISSN 0003-1488. URL http://www.ncbi.nlm.nih.gov/pubmed/11110459. 10

R. D. Kealy, D. F. Lawler, J. M. Ballam, S. L. Mantz, D. N. Biery, E. H. Greeley, G. Lust, M. Segre, G. K. Smith, and H. D. Stowe. Effects of diet restriction on life span and age-related changes in dogs. *Journal of the American Veterinary Medical Association*, 220(9):1315–1320, 2002. URL http://avmajournals.avma.org/doi/abs/10.2460/javma.2002.220.1315. 9

Kennel Club. General Puppy Health: Vaccinations, Fleas & Health Checks, 2015. URL http://www.thekennelclub.org.uk/getting-a-dog-or-puppy/general-advice-about-caring-for-your-new-puppy-or-dog/general-puppy-health/. 183

E. F. Kirkness, V. Bafna, A. L. Halpern, S. Levy, K. Remington, D. B. Rusch, A. L. Delcher, M. Pop, W. Wang, C. M. Fraser, and J. C. Venter. The dog genome: survey sequencing and comparative analysis. *Science*, 301(5641):1898–1903, Sept. 2003. ISSN 1095-9203. doi: 10.1126/science.1086432. URL http://www.ncbi.nlm.nih.gov/pubmed/14512627. 14

C. M. Knottenbelt, S. Bawazeer, J. Hammond, D. Mellor, and D. G. Watson. Nicotine hair concentrations in dogs exposed to environmental tobacco smoke: a pilot study. *The Journal of small animal practice*, 53(11):623–626, Oct. 2012. ISSN 1748-5827. doi: 10.1111/j.1748-5827.2012.01284.x. URL http://www.ncbi.nlm.nih.gov/pubmed/23020087. 72, 77

R. I. Krontveit, A. Nødtvedt, B. K. Sævik, E. Ropstad, H. K. Skogmo, and C. Trangerud. A prospective study on canine hip dysplasia and growth in a cohort of four large breeds in Norway (1998-2001). *Preventive Veterinary Medicine*, 97(3-4):252–263, Dec. 2010. ISSN 1873-1716. doi: 10.1016/j.prevetmed.2010.09.015. URL http://www.ncbi.nlm.nih.gov/pubmed/20956024. 13, 243

R. I. Krontveit, A. Nødtvedt, B. K. Sævik, E. Ropstad, and C. Trangerud. Housing- and exercise-related risk factors associated with the development of hip dysplasia as determined by radiographic evaluation in a prospective cohort of Newfoundland, Labrador Retrievers, Leonbergers, and Irish Wolfhounds in Norway. *American journal of public health*, 73(6):838–846, 2012. 188

M. Last and M. B. Adelaide. The Iceberg “Completing The Clinical Picture” In General Practice. *The Lancet*, 2(7297):28–31, 1963. doi: 10.1016/S0140-6736(63)92546-7. 179, 180

D. F. Lawler, R. H. Evans, B. T. Larson, E. L. Spitznagel, M. R. Ellersieck, and R. D. Kealy. Influence of lifetime food restriction on causes, time, and predictors of death in dogs. *Journal of the American Veterinary Medical Association*, 226(2):225–231, Jan. 2005. ISSN 0003-1488. URL http://www.ncbi.nlm.nih.gov/pubmed/15706972. 9

K. Leffondre, M. Abrahamowicz, J. Siemiatycki, and B. Rachet. Modeling Smoking History: A Comparison of Different Approaches. *American Journal of Epidemiology*, 156(9):813–823, Nov. 2002. ISSN 00029262. doi: 10.1093/aje/kwf122. URL http://aje.oupjournals.org/cgi/doi/10.1093/aje/kwf122. 37

S. J. Lewis, N. M. Cherry, R. McL Niven, P. V. Barber, K. Wilde, and A. C. Povey. Cotinine levels and self-reported smoking status in patients attending a bronchoscopy clinic. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*, 8(3-4):218–228, 2003. ISSN 1354-750X. doi: 10.1080/1354750031000120125. URL http://www.ncbi.nlm.nih.gov/pubmed/12944174. 72

L. Lin. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Bio-*

metrics, 45(1):255–268, 1989. URL http://www.jstor.org/stable/10.2307/2532051. 41, 56, 58

K. Lindblad-Toh, C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe, M. Kamal, M. Clamp, J. L. Chang, E. J. Kulbokas, M. C. Zody, E. Mauceli, X. Xie, M. Breen, R. K. Wayne, E. A. Ostrander, C. P. Ponting, F. Galibert, D. R. Smith, P. J. DeJong, E. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C.-W. Chin, A. Cook, J. Cuff, M. J. Daly, D. DeCaprio, S. Gnerre, M. Grabherr, M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L. Kim, K.-P. Koepfli, H. G. Parker, J. P. Pollinger, S. M. J. Searle, N. B. Sutter, R. Thomas, C. Webber, J. Baldwin, A. Abebe, A. Abouelleil, L. Aftuck, M. Ait-Zahra, T. Aldredge, N. Allen, P. An, S. Anderson, C. Antoine, H. Arachchi, A. Aslam, L. Ayotte, P. Bachantsang, A. Barry, T. Bayul, M. Benamara, A. Berlin, D. Bessette, B. Blitshteyn, T. Bloom, J. Blye, L. Boguslavskiy, C. Bonnet, B. Boukhgalter, A. Brown, P. Cahill, N. Calixte, J. Camarata, Y. Cheshatsang, J. Chu, M. Citroen, A. Collymore, P. Cooke, T. Dawoe, R. Daza, K. Decktor, S. DeGray, N. Dhargay, K. Dooley, K. Dooley, P. Dorje, K. Dorjee, L. Dorris, N. Duffey, A. Dupes, O. Egbiremolen, R. Elong, J. Falk, A. Farina, S. Faro, D. Ferguson, P. Ferreira, S. Fisher, M. FitzGerald, K. Foley, C. Foley, A. Franke, D. Friedrich, D. Gage, M. Garber, G. Gearin, G. Giannoukos, T. Goode, A. Goyette, J. Graham, E. Grandbois, K. Gyaltsen, N. Hafez, D. Hagopian, B. Hagos, J. Hall, C. Healy, R. Hegarty, T. Honan, A. Horn, N. Houde, L. Hughes, L. Hunnicutt, M. Husby, B. Jester, C. Jones, A. Kamat, B. Kanga, C. Kells, D. Khazanovich, A. C. Kieu, P. Kisner, M. Kumar, K. Lance, T. Landers, M. Lara, W. Lee, J.-P. Leger, N. Lennon, L. Leuper, S. LeVine, J. Liu, X. Liu, Y. Lokyitsang, T. Lokyitsang, A. Lui, J. Macdonald, J. Major, R. Marabella, K. Maru, C. Matthews, S. McDonough, T. Mehta, J. Meldrim, A. Melnikov, L. Meneus, A. Mihalev, T. Mihova, K. Miller, R. Mittelman, V. Mlenga, L. Mulrain, G. Munson, A. Navidi, J. Naylor, T. Nguyen, N. Nguyen, C. Nguyen, T. Nguyen, R. Nicol, N. Norbu, C. Norbu, N. Novod, T. Nyima, P. Olandt, B. O'Neill, K. O'Neill, S. Osman, L. Oyono, C. Patti, D. Perrin, P. Phunkhang, F. Pierre, M. Priest, A. Rachupka, S. Raghuraman, R. Rameau, V. Ray, C. Raymond, F. Rege, C. Rise, J. Rogers, P. Rogov, J. Sahalie, S. Settipalli, T. Sharpe, T. Shea, M. Sheehan, N. Sherpa, J. Shi, D. Shih, J. Sloan, C. Smith, T. Sparrow, J. Stalker, N. Stange-Thomann,

S. Stavropoulos, C. Stone, S. Stone, S. Sykes, P. Tchuinga, P. Tenzing, S. Tesfaye, D. Thoulutsang, Y. Thoulutsang, K. Topham, I. Topping, T. Tsamla, H. Vassiliev, V. Venkataraman, A. Vo, T. Wangchuk, T. Wangdi, M. Weiand, J. Wilkinson, A. Wilson, S. Yadav, S. Yang, X. Yang, G. Young, Q. Yu, J. Zainoun, L. Zembek, A. Zimmer, and E. S. Lander. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438 (7069):803–19, Dec. 2005. ISSN 1476-4687. doi: 10.1038/nature04338. URL http://www.ncbi.nlm.nih.gov/pubmed/16341006. 14

A. J. Littman, E. J. Boyko, I. G. Jacobson, J. Horton, G. D. Gackstetter, B. Smith, T. Hooper, T. S. Wells, P. J. Amoroso, and T. C. Smith. Assessing nonresponse bias at follow-up in a large prospective cohort of relatively young and mobile military service members. *BMC medical research methodology*, 10:99, Jan. 2010. ISSN 1471-2288. doi: 10.1186/1471-2288-10-99. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2984503&tool=pmcentrez&rendertype=abstract. 71

S. S. Mahmood, D. Levy, R. S. Vasan, and T. J. Wang. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet*, 383(9921):999–1008, Sept. 2013. ISSN 1474-547X. doi: 10.1016/S0140-6736(13)61752-3. URL http://www.ncbi.nlm.nih.gov/pubmed/24084292. 2

G. B. McBride. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. Technical report, National Institute of Water and Atmospheric Research Ltd., Hamilton, 2005. 41, 56, 58

C. S. McBrien, A. Vezzoni, and M. G. Conzemius. Growth dynamics of the canine proximal tibial physis. *Veterinary surgery : VS*, 40(4):389–94, June 2011. ISSN 1532-950X. doi: 10.1111/j.1532-950X.2011.00825.x. URL http://www.ncbi.nlm.nih.gov/pubmed/21470252. 76

E. L. Melanson and P. S. Freedson. Physical activity assessment: A review of methods. *Critical reviews in food science and nutrition*, 36(5):385–396, 1996. ISSN 1040-8398. doi: 10.1080/10408399609527732. 187

Met Office. Climate summary by region, 2013, 2014. URL http:

//www.metoffice.gov.uk/climate/uk/summaries/2013/annual/regional-values. 139

K. E. Michel and D. C. Brown. Determination and application of cut points for accelerometer-based activity counts of activities with differing intensity in pet dogs. *American Journal of Veterinary Research*, 72(7):866–870, 2011. 193, 198, 213, 214

G. Mooney, M. Krakowiecki, and D. Trunzo. Does Prefilling Responses on a Longitudinal Establishment Survey Stem Sample Attrition? In *American Association For Public Opinion Research*, pages 5917–5924, 2009. URL https://www.amstat.org/sections/srms/Proceedings/y2009/Files/400048.pdf. 78

E. D. Moreira, V. M. M. de Souza, M. Sreenivasan, N. L. Lopes, R. B. Barreto, and L. P. de Carvalho. Peridomestic risk factors for canine leishmaniasis in urban dwellings: new findings from a prospective study in Brazil. *The American Journal of Tropical Medicine and Hygiene*, 69(4):393–397, Oct. 2003. ISSN 0002-9637. URL http://www.ncbi.nlm.nih.gov/pubmed/14640499. 9

E. D. Moreira, V. M. Mendes de Souza, M. Sreenivasan, E. G. Nascimento, and L. Pontes de Carvalho. Assessment of an optimized dog-culling program in the dynamics of canine Leishmania transmission. *Veterinary Parasitology*, 122(4):245–252, Aug. 2004. ISSN 0304-4017. doi: 10.1016/j.vetpar.2004.05.019. URL http://www.ncbi.nlm.nih.gov/pubmed/15262002. 9

R. Morrison, V. Penpraze, R. Greening, T. Underwood, J. J. Reilly, and P. S. Yam. Correlates of objectively measured physical activity in dogs. *The Veterinary Journal*, 199(2):120–126, 2014. doi: 10.1016/j.tvjl.2013.11.023. URL http://dx.doi.org/10.1016/j.tvjl.2013.11.023. 213, 215

J. K. Murray, W. J. Browne, M. A. Roberts, A. Whitmarsh, and T. J. Gruffydd-Jones. Number and ownership profiles of cats and dogs in the UK. *The Veterinary record*, 166(6):163–8, Feb. 2010. ISSN 2042-7670. doi: 10.1136/vr.b4712. URL http://www.ncbi.nlm.nih.gov/pubmed/20139379. 18

J. K. Murray, R. A. Casey, E. Gale, C. A. T. Buffington, and T. J. Gruffydd-Jones. The Bristol Cats Study : Recruitment and characteristics of the cohort, 2015. URL http://www.bristol.ac.uk/media-library/sites/vetscience/documents/svemp2015.pdf. 243

National Records of Scotland. Scotland Census 2011: Table QS102SC. Technical report, National Records of Scotland, 2011. 148

Nature Editors. Rise of the citizen scientist. *Nature*, 524(7565):265, 2015. doi: 10.1038/524265a. URL http://www.nature.com/news/rise-of-the-citizen-scientist-1.18192. 254

J. Neter and J. Waksberg. Three-month recall loss and reference period effects. In *Response errors in collection of expenditures data by household interviews: An experimental study, Volume 3*, chapter 7, pages 51–65. U.S. Department of Commerce, Bureau of the Census, 1965, 1965. 80

A. Nielen. *Genetic-epidemiologic research in a birth cohort of boxer dogs*. PhD thesis, Universiteit Utrecht, 2000. 243

A. Nielen, I. van der Gaag, B. Knol, and Y. Schukken. Investigation of mortality and pathological changes in a 14-month birth cohort of boxer puppies. *The Veterinary Record*, 142(22):602–606, May 1998. ISSN 0042-4900. URL http://www.ncbi.nlm.nih.gov/pubmed/9682419. 12, 13

A. Nielen, L. Janss, and B. Knol. Heritability estimations for diseases, coat color, body weight, and height in a birth cohort of Boxers. *American Journal of Veterinary Research*, 62(8):1198–1206, 2001. URL http://avmajournals.avma.org/doi/abs/10.2460/ajvr.2001.62.1198. 12

A. Nielen, B. Knol, M. van Hagen, and I. van der Gaag. Genetic and epidemiological investigation of a birth cohort of boxers. *Tijdschr Diergeneeskd*, 128 (19):586–90, 2003. 13

NINFEA. NINFEA, 2014. URL https://www.progettoninfea.it/index_en. 6

A. Nødtvedt, J. Guitian, A. Egenvall, U. Emanuelson, and D. U. Pfeiffer. The spatial distribution of atopic dermatitis cases in a population of insured Swedish dogs. *Preventive Veterinary Medicine*, 78(3-4):210–222, Mar. 2007. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2006.10.007. URL http://www.ncbi.nlm.nih.gov/pubmed/17112611. 7

Office for Naitonal Statistics. England and Wales Census 2011: Table KS01. Technical report, Office for National Statistics, 2011. 148

G. Oliva, A. Scalone, V. Foglia Manzillo, M. Gramiccia, A. Pagano, T. Di Muccio, and L. Gradoni. Incidence and Time Course of Leishmania infantum Infections Examined by Parasitological , Serologic , and Nested-PCR Techniques in a Cohort of NaÄśve Dogs Exposed to Three Consecutive Transmission Seasons. *Journal of Clinical Microbiology*, 44(4):1318–1322, 2006. doi: 10.1128/JCM.44.4.1318. URL http://jcm.asm.org/content/44/4/1318.short. 9

D. G. O'Neill, D. B. Church, P. D. McGreevy, P. C. Thomson, and D. C. Brodbelt. Approaches to Canine Health Surveillance. *Canine Genetics and Epidemiology*, 1(2), 2014a. 7

D. G. O'Neill, D. B. Church, P. D. McGreevy, P. C. Thomson, and D. C. Brodbelt. Prevalence of disorders recorded in dogs attending primary-care veterinary practices in England. *PloS one*, 9(3): e90501, Jan. 2014b. ISSN 1932-6203. doi: 10.1371/journal.pone.0090501. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3942437&tool=pmcentrez&rendertype=abstract. 139, 179

D. Otranto, G. Testini, F. Dantas-Torres, M. S. Latrofa, P. P. V. d. P. Diniz, D. de Caprariis, R. P. Lia, N. Mencke, D. Stanneck, G. Capelli, and E. B. Breitschwerdt. Diagnosis of canine vector-borne diseases in young dogs: a longitudinal study. *Journal of Clinical Microbiology*, 48(9):3316–3324, Sept. 2010. ISSN 1098-660X. doi: 10.1128/JCM.00379-10. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2937705&tool=pmcentrez&rendertype=abstract. 9

C. Overy, L. Reynolds, and E. Tansey. *History of Avon Longitudinal Study of Parents and Children (ALSPAC), c.1980-2000. Wellcome Witnesses to Twentieth Century Medicine, vol. 44.* Queen Mary University of London, London, 2012. 4, 11

M. Paranhos-Silva, E. Nascimento, M. Melro, G. Oliveira, W. dos Santos, L. Pontes-de Carvalho, and A. Oliveira-dos Santos. Cohort study on canine emigration and Leishmania infection in an endemic area for American visceral leishmaniasis. Implications for the disease control. *Acta Tropica*, 69(1):75–83, Mar. 1998. ISSN 0001-706X. URL http://www.ncbi.nlm.nih.gov/pubmed/9588243. 9

K. Passaro, R. Little, D. Savitz, and J. Noss. The effect of maternal drinking before conception and in early pregnancy on infant birthweight. *Epidemiology*, 7(4):377–383, 1996. URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+effect+of+maternal+drinking+before+conception+and+in+early+pregnancy+on+infant+birthweight.#0. 4

D. L. Patrick, A. Cheadle, D. C. Thompson, P. Diehr, T. Koepsell, and S. Kinne. The validity of self-reported smoking: a review and meta-analysis. *American journal of public health*, 84(7):1086–93, July 1994. ISSN 0090-0036. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1614767&tool=pmcentrez&rendertype=abstract. 72

Pet Food Manufacturers Association. Pet Population 2014, 2014. URL http://www.pfma.org.uk/pet-population-2014/. 18

J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R. D. C. Team. nlme: Linear and Nonlinear Mixed Effects Models, 2013. URL http://cran.r-project.org/package=nlme. 83, 193, 221

M. Plummer and A. Stukalov. rjags: Bayesian graphical models using MCMC, 2014. URL http://mcmc-jags.sourceforge.net. 87

M. Plummer, N. Best, K. Cowles, and K. Vines. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11, 2006. 87

M. Powers, D. Biery, D. Lawler, R. Evans, F. Shofer, P. Mayhew, T. Gregor, R. Kealy, and G. Smith. Use of the caudolateral curvilinear osteophyte as an early marker for future development of osteoarthritis associated with hip dysplasia in dogs. *Journal of the American Veterinary Medical Association*, 225(2):233–237, 2004. URL https://avmajournals.avma.org/doi/abs/10.2460/javma.2004.225.233. 10

C. A. Pugh, B. M. C. Bronsvoort, I. G. Handel, K. M. Summers, and D. N. Clements. What can cohort studies in the dog tell us? *Canine Genetics and Epidemiology*, 1(1):5, 2014. ISSN 2052-6687. doi: 10.1186/2052-6687-1-5. URL http://www.cgejournal.org/content/1/1/5/abstract. 1

C. A. Pugh, B. M. C. Bronsvoort, I. G. Handel, K. M. Summers, and D. N. Clements. Dogslife: A cohort study of Labrador Retriever health. In M. Brennan, editor, *Society for Veterinary Epidemiology and Preventive Medicine*, pages 1–11, Ghent, 2015a. 150, 250

C. A. Pugh, B. M. C. Bronsvoort, I. G. Handel, K. M. Summers, and D. N. Clements. Dogslife: A cohort study of Labrador Retrievers in the UK. *Preventive veterinary medicine*, 2015b. doi: 10.1016/j.prevetmed.2015.06.020. URL http://www.sciencedirect.com/science/article/pii/S0167587715002329. 17, 83, 250

C. A. Pugh, K. M. Summers, B. M. C. Bronsvoort, I. G. Handel, and D. N. Clements. Validity of Internet-Based Longitudinal Study Data: The Elephant in the Virtual Room. *Journal of Medical Internet Research*, 17(4):e96, 2015c. ISSN 1438-8871. doi: 10.2196/jmir.3530. URL http://www.jmir.org/2015/4/e96/. 38

S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3):559–575, 2007. ISSN 00029297. doi: 10.1086/519795. 221

R. J. Quinnell, O. Courtenay, L. M. Garcez, P. M. Kaye, M.-A. Shaw, C. Dye, and M. J. Day. IgG subclass responses in a longitudinal study of canine visceral leishmaniasis. *Veterinary Immunology and Immunopathology*, 91(3-4):161–168, Feb. 2003. ISSN 01652427. doi: 10.1016/S0165-2427(02)00311-2. URL http://linkinghub.elsevier.com/retrieve/pii/S0165242702003112. 9

R. J. Quinnell, C. Carson, R. Reithinger, L. M. Garcez, and O. Courtenay. Evaluation of rK39 rapid diagnostic tests for canine visceral leishmaniasis: longitudinal study and meta-analysis. *PLoS Neglected Tropical Diseases*, 7(1):e1992, Jan. 2013. ISSN 1935-2735. doi: 10.1371/journal.pntd.0001992. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3542187&tool=pmcentrez&rendertype=abstract. 9

R Core Team. R: A language and environment for statistical computing., 2013. URL http://www.r-project.org/. 23, 144, 220

L. S. Radloff. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1(3):385–401, June 1977. ISSN 0146-6216. doi: 10.1177/014662167700100306. URL http://apm.sagepub.com/cgi/doi/10.1177/014662167700100306. 37

S. A. Reijneveld and K. Stronks. The impact of response bias on estimates of health care utilization in a metropolitan area: the use of administrative data. *International journal of epidemiology*, 28(6):1134–40, Dec. 1999. ISSN 0300-5771. URL http://www.ncbi.nlm.nih.gov/pubmed/10661659. 71

E. Repapi, I. Sayers, L. V. Wain, P. R. Burton, T. Johnson, M. Obeidat, J. H. Zhao, A. Ramasamy, G. Zhai, V. Vitart, J. E. Huffman, W. Igl, E. Albrecht, P. Deloukas, J. Henderson, R. Granell, W. L. McArdle, A. R. Rudnicka, I. Barroso, R. J. F. Loos, N. J. Wareham, L. Mustelin, T. Rantanen, I. Surakka, M. Imboden, H. E. Wichmann, I. Grkovic, S. Jankovic, L. Zgaga, A.-L. Hartikainen, L. Peltonen, U. Gyllensten, A. Johansson, G. Zaboli, H. Campbell, S. H. Wild, J. F. Wilson, S. Gläser, G. Homuth, H. Völzke, M. Mangino, N. Soranzo, T. D. Spector, O. Polasek, I. Rudan, A. F. Wright, M. Heliövaara, S. Ripatti, A. Pouta, A. T. Naluai, A.-C. Olin, K. Torén, M. N. Cooper, A. L. James, L. J. Palmer, A. D. Hingorani, S. G. Wannamethee, P. H. Whincup, G. D. Smith, S. Ebrahim, T. M. McKeever, I. D. Pavord, A. K. MacLeod, A. D. Morris, D. J. Porteous, C. Cooper, E. Dennison, S. Shaheen, S. Karrasch, E. Schnabel, H. Schulz, H. Grallert, N. Bouatia-Naji, J. Delplanque, P. Froguel, J. D. Blakey, J. R. Britton, R. W. Morris, J. W. Holloway, D. A. Lawlor, J. Hui, F. Nyberg, M.-R. Jarvelin, C. Jackson, M. Kähönen, J. Kaprio, N. M. Probst-Hensch, B. Koch, C. Hayward, D. M. Evans, P. Elliott, D. P. Strachan, I. P. Hall, and M. D. Tobin. Genome-wide association study identifies five loci associated with lung function. *Nature genetics*, 42(1):36–44, 2010. ISSN 1061-4036. doi: 10.1038/ng.501. 15

L. Richiardi, I. Baussano, L. Vizzini, J. Douwes, N. Pearce, and F. Merletti. Feasibility of recruiting a birth cohort through the Internet : the experience of the NINFEA cohort. *European Journal of Epidemiology*, 22(12):831–837, 2007. doi: 10.1007/sl0654-007-9194-2. 4

J. J. Runge, D. N. Biery, D. F. Lawler, T. P. Gregor, R. H. Evans, R. D. Kealy, S. D. Szabo, and G. K. Smith. The effects of lifetime food restriction on the development of osteoarthritis in the canine shoulder. *Veterinary Surgery*, 37(1): 102–107, Jan. 2008. ISSN 1532-950X. doi: 10.1111/j.1532-950X.2007.00354.x. URL http://www.ncbi.nlm.nih.gov/pubmed/18199063. 10

T. Russell, M. Crawford, and L. Woodby. Measurements for active cigarette smoke exposure in prevalence and cessation studies: why simply asking pregnant women isn't enough. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*, 6(Suppl 2):S141–S151, Apr. 2004. ISSN 1462-2203. doi: 10.1080/14622200410001669141. URL http://www.ncbi.nlm.nih.gov/pubmed/15203817. 72

B. K. Sævik, E. M. Skancke, and C. Trangerud. A longitudinal study on diarrhoea and vomiting in young dogs of four large breeds. *Acta Veterinaria Scandinavica*, 54(1):8–16, Jan. 2012. ISSN 1751-0147. doi: 10.1186/1751-0147-54-8. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3293024&tool=pmcentrez&rendertype=abstract. 13, 183, 184, 185, 243

M. D. Salman, J. Hutchison, R. Ruch-Gallie, L. Kogan, J. C. New, P. H. Kass, and J. M. Scarlett. Behavioral Reasons for Relinquishment of Dogs and Cats to 12 Shelters. *Journal of Applied Animal Welfare Science*, 3(2):93–106, 2000. ISSN 1088-8705. doi: 10.1207/S15327604JAWS0302_2. 181

O. J. G. Schiepers, S. E. Harris, A. J. Gow, A. Pattie, C. E. Brett, J. M. Starr, and I. J. Deary. APOE E4 status predicts age-related cognitive decline in the ninth decade: longitudinal follow-up of the Lothian Birth Cohort 1921. *Molecular psychiatry*, 17(3):315–324, Mar. 2012. ISSN 1476-5578. doi: 10.1038/mp.2010.137. URL http://www.ncbi.nlm.nih.gov/pubmed/21263443. 6

W.-P. Schmidt, S. P. Luby, B. Genser, M. L. Barreto, and T. Clasen. Estimating the longitudinal prevalence of diarrhea and other episodic diseases: continuous versus intermittent surveillance. *Epidemiology*, 18(5):537–543, Sept. 2007. ISSN 1044-3983. doi: 10.1097/EDE.0b013e318093f3ce. URL http://www.ncbi.nlm.nih.gov/pubmed/17603390. 81

S. M. Schmutz and T. G. Berryere. Genes affecting coat colour and pattern in domestic dogs: a review. *Animal genetics*, 38(6):539–549, Dec. 2007. ISSN 1365-2052. doi: 10.1111/j.1365-2052.2007.01664.x. URL http://www.ncbi.nlm.nih.gov/pubmed/18052939. 84

A. Shillington and M. Reed. Testing the Length of Time Theory of Recall Decay: Examining Substance Use Report Stability With 10 Years of National Longitudinal Survey of Youth Data. *Substance use and misuse*, 46(9): 1105–1112, 2011. doi: 10.3109/10826084.2010.548436.Testing. URL http://informahealthcare.com/doi/abs/10.3109/10826084.2010.548436. 80

S. Sieri, A. Agudo, E. Kesse, K. Klipstein-Grobusch, B. San-José, A. A. Welch, V. Krogh, R. Luben, N. Allen, K. Overvad, A. Tjønneland, F. Clavel-Chapelon, A. Thiébaut, a. B. Miller, H. Boeing, M. Kolyva, C. Saieva, E. Celentano, M. C. Ocké, P. H. M. Peeters, M. Brustad, M. Kumle, M. Dorronsoro, A. Fernandez Feito, I. Mattisson, L. Weinehall, E. Riboli, and N. Slimani. Patterns of alcohol consumption in 10 European countries participating in the European Prospective Investigation into Cancer and Nutrition (EPIC) project. *Public health nutrition*, 5(6B):1287–96, Dec. 2002. ISSN 1368-9800. doi: 10.1079/PHN2002405. URL http://www.ncbi.nlm.nih.gov/pubmed/12639233. 37

B. Smith, T. C. Smith, G. C. Gray, and M. A. K. Ryan. When epidemiology meets the Internet: Web-based surveys in the Millennium Cohort Study. *American Journal of Epidemiology*, 166(11):1345–1354, Dec. 2007. ISSN 1476-6256. doi: 10.1093/aje/kwm212. URL http://www.ncbi.nlm.nih.gov/pubmed/17728269. 6

B. H. Smith, K. I. Penny, A. M. Purves, C. Munro, B. Wilson, J. Grimshaw, W. A. Chambers, and W. C. Smith. The Chronic Pain Grade questionnaire: validation and reliability in postal research. *Pain*, 71(2):141–147, June 1997. ISSN 0304-3959. URL http://www.ncbi.nlm.nih.gov/pubmed/9211475. 37

G. K. Smith, E. R. Paster, M. Y. Powers, D. F. Lawler, D. N. Biery, F. S. Shofer, P. J. McKelvie, and R. D. Kealy. Lifelong diet restriction and radiographic evidence of osteoarthritis of the hip joint in dogs. *Journal of the American Veterinary Medical Association*, 229(5):690–693, 2006. 10, 140

G. K. Smith, D. F. Lawler, D. N. Biery, M. Y. Powers, F. Shofer, T. P. Gregor, G. T. Karbe, M. B. McDonald-Lynch, R. H. Evans, and R. D. Kealy. Chronology of hip dysplasia development in a cohort of 48 Labrador retrievers followed for life. *Veterinary Surgery*, 41(1):20–33, Jan. 2012. ISSN 1532-

950X. doi: 10.1111/j.1532-950X.2011.00935.x. URL http://www.ncbi.nlm.nih.gov/pubmed/23253036. 10

A. J. Søgaard, R. Selmer, E. Bjertness, and D. Thelle. The Oslo Health Study: The impact of self-selection in a large, population-based survey. *International Journal for Equity in Health*, 3(3):1–12, 2004. doi: 10.1186/1475-9276-3-3. URL http://www.equityhealthj.com/content/3/1/3. 35, 242

J. Stavisky, A. D. Radford, R. Gaskell, S. Dawson, A. German, B. Parsons, S. Clegg, J. Newman, and G. Pinchbeck. A case-control study of pathogen and lifestyle risk factors for diarrhoea in dogs. *Preventive Veterinary Medicine*, 99 (2-4):185–192, 2011. ISSN 01675877. doi: 10.1016/j.prevetmed.2011.02.009. URL http://dx.doi.org/10.1016/j.prevetmed.2011.02.009. 184

J. Steiss, K. Braund, J. Wright, S. Lenz, J. Hudson, W. Brawner, J. Hathcock, R. Purohit, L. Bell, and R. Horne. Coccygeal Muscle Injury in English Pointers (Limber Tail). *Journal of veterinary internal medicine*, 13:540–548, 1999. 217, 239

J. E. Steiss. Limber Tail Syndrome in Hunting Dogs. *American Canine Sports Medicine Association Newsletter*, 2(3), 1996. 218

J. E. Steiss. What is limber tail syndrome? *Field Trial Magazine*, 1(1):12–13, 1997a. 218

J. E. Steiss. Frozen tail or limber tail in working dogs. *Veterinary Record*, 141 (7):179, 1997b. 217

J. E. Steiss. What is limber tail syndrome? *Canine Practice*, 22(1):5–6, 1997c. 218

J. E. Steiss. Muscle disorders and rehabilitation in canine athletes. *The Veterinary Clinics of North America: Small Animal Practice*, 32(1):267–285, 2002. 217

J. E. Steiss and J. C. Wright. Limber Tail Syndrome in Hunting Dogs. *Sports Medicine Program Newsletter, Auburn Univeristy College of Veterinary Medicine*, 1995. 218

E. Stergiakouli, R. Gaillard, J. M. Tavaré, N. Balthasar, R. J. Loos, H. R. Taal, D. M. Evans, F. Rivadeneira, B. St Pourcain, A. G. Uitterlinden, J. P. Kemp,

A. Hofman, S. M. Ring, T. J. Cole, V. W. Jaddoe, G. Davey Smith, and N. J. Timpson. Genome-wide association study of height-adjusted BMI in childhood identifies functional variant in ADCY3. *Obesity*, 22(10):2252–2259, 2014. ISSN 19307381. doi: 10.1002/oby.20840. URL http://doi.wiley.com/10.1002/oby.20840. 15

P. S. Sullivan, C. M. Khosropour, N. Luisi, M. Amsden, T. Coggia, G. M. Wingood, and R. J. DiClemente. Bias in Online recruitment and retention of racial and ethnic minority men who have sex with men. *Journal of Medical Internet Research*, 13(2), 2011. ISSN 14388871. doi: 10.2196/jmir.1797. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3221372/. 242

N. B. Sutter, D. S. Mosher, M. M. Gray, and E. A. Ostrander. Morphometrics within dog breeds are highly reproducible and dispute Rensch's rule. *Mammalian genome*, 19(10-12):713–23, 2008. ISSN 0938-8990. doi: 10.1007/s00335-008-9153-6. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2748280&tool=pmcentrez&rendertype=abstract. 140

S. D. Szabo, D. N. Biery, D. F. Lawler, F. S. Shofer, M. Y. Powers, R. D. Kealy, and G. K. Smith. Evaluation of a circumferential femoral head osteophyte as an early indicator of osteoarthritis characteristic of canine hip dysplasia in dogs. *Journal of the American Veterinary Medical Association*, 231(6):889–892, 2007. 10

The Kennel Club. Labrador Retriever Breed Standard, 2014. URL http://www.thekennelclub.org.uk/services/public/breed/standard.aspx?id=2048. 87, 140

The Million Women Collaborative Study Group. The Million Women Study: design and characteristics of the study population. *Breast Cancer Research*, 1(1):73–80, Jan. 1999. ISSN 1465-5411. doi: 10.1186/bcr16. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=13913&tool=pmcentrez&rendertype=abstract. 5

The Million Women Collaborative Study Group. Breast cancer and hormone-replacement therapy: the Million Women Study. *Lancet*, 362(9392):419–427, Oct. 2003. ISSN 1474-547X. doi: 10.1016/S0140-6736(03)14598-9. URL http://www.ncbi.nlm.nih.gov/pubmed/22262621. 5

T. M. Therneau. A Package for Survival Analysis in S, 2014. URL http://cran.r-project.org/package=survival. 24, 147

M. Thrusfield. Observational Studies. In *Veterinary Epidemiology*, chapter 15, pages 266–288. Blackwell Science Ltd, Oxford, 3rd edition, 2005. ISBN 978-0-632-06397-0. 1

M. V. Thrusfield, P. E. Holt, and R. H. Muirhead. Acquired urinary incontinence in bitches: its incidence and relationship to neutering practices. *The Journal of small animal practice*, 39(12):559–566, Dec. 1998. ISSN 0022-4510. URL http://www.ncbi.nlm.nih.gov/pubmed/9888109. 11

V. B. L. Torman and S. A. Camey. Bayesian models as a unified approach to estimate relative risk (or prevalence ratio) in binary and polytomous outcomes. *Emerging Themes in Epidemiology*, 12(8), 2015. ISSN 1742-7622. doi: 10.1186/s12982-015-0030-y. URL http://www.ete-online.com/content/12/1/8. 252

US Department of Health, Education and Welfare. US National Health Survey. Health interview responses compared with medical records. Public health service publication, series D, no. 5. Technical report, US Government, Washington DC, 1961. 182

S. van der Beek, A. Nielen, Y. Schukken, and E. Brascamp. Evaluation of genetic, common-litter, and within-litter effects on preweaning mortality in a birth cohort of puppies. *American Journal of Veterinary Research*, 60(9):1106–1110, Sept. 1999. ISSN 0002-9645. URL http://www.ncbi.nlm.nih.gov/pubmed/10490080. 12, 13, 14

M. van Hagen, B. Ducro, J. van den Broek, and B. Knol. Incidence, risk factors, and heritability estimates of hind limb lameness caused by hip dysplasia in a birth cohort of Boxers. *American Journal of Veterinary Research*, 66(2):307–312, 2005a. URL http://avmajournals.avma.org/doi/pdf/10.2460/ajvr.2005.66.307. 12

M. van Hagen, B. Ducro, J. van den Broek, and B. Knol. Life expectancy in a birth cohort of Boxers followed up from weaning to 10 years of age. *American Journal* *of Veterinary Research*, 66(9):1646–1650, 2005b. URL http://avmajournals.avma.org/doi/abs/10.2460/ajvr.2005.66.1646. 13

VetCompass. VetCompass, 2014. URL http://www.rvc.ac.uk/VetCompass/Index.cfm. 246

N. M. Warrington, L. D. Howe, L. Paternoster, M. Kaakinen, S. Herrala, V. Huikari, Y. Y. Wu, J. P. Kemp, N. J. Timpson, B. S. Pourcain, G. Davey Smith, K. Tilling, M.-R. Jarvelin, C. E. Pennell, D. M. Evans, D. A. Lawlor, L. Briollais, and L. J. Palmer. A genome-wide association study of body mass index across early life and childhood. *International Journal of Epidemiology*, pages 1–13, 2015. ISSN 0300-5771. doi: 10.1093/ije/dyv077. URL http://www.ije.oxfordjournals.org/cgi/doi/10.1093/ije/dyv077. 15

A.-K. Wejdmark, B. Bonnett, A. Hedhammar, and T. Fall. Lifestyle risk factors for progesterone-related diabetes mellitus in elkhounds - a case-control study. *The Journal of small animal practice*, 52(5):240–5, May 2011. ISSN 1748-5827. doi: 10.1111/j.1748-5827.2011.01052.x. URL http://www.ncbi.nlm.nih.gov/pubmed/21539568. 140

J. Wheeler, D. Sethi, J. Cowden, P. Wall, L. Rodrigues, D. Tompkins, M. Hudson, and P. Roderick. Study of infectious intestinal disease in England : rates in the community, presenting to general practice, and reported to national surveillance. *BMJ*, 318(April):1846–1850, 1999. URL http://www.bmj.com/content/318/7190/1046. 79

C. M. Wilkins. Frozen tail or limber tail in working dogs. *Veterinary Record*, 140:588, 1997. 217

D. Wolke, A. Waylen, M. Samara, C. Steer, R. Goodman, T. Ford, and K. Lamberts. Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *British Journal of Psychiatry*, 195(3):249–256, 2009. ISSN 00071250. doi: 10.1192/bjp.bp.108.053751. 242

World Health Organisation. WORLD HEALTH REPORT 2002: Reducing Risks, Promoting Healthy Life. Technical report, World Health Organisation, Geneva, 2002. URL http://www.who.int/whr/2002/en/. 212

D. J. Wrigglesworth, E. S. Mort, S. L. Upton, and A. T. Miller. Accuracy of the use of triaxial accelerometry for measuring daily activity as a predictor of daily maintenance energy requirement in healthy adult Labrador Retrievers. *American Journal of Veterinary Research*, 72(9):1151–1155, 2011. ISSN 00029645. doi: 10.2460/ajvr.72.9.1151. 213

P. S. Yam, V. Penpraze, D. Young, M. S. Todd, A. D. Cloney, K. A. Houston-Callaghan, and J. J. Reilly. Validity, practical utility and reliability of Actigraph accelerometry for the measurement of habitual physical activity in dogs. *The Journal of small animal practice*, 52(2):86–91, Feb. 2011. ISSN 1748-5827. doi: 10.1111/j.1748-5827.2010.01025.x. URL http://www.ncbi.nlm.nih.gov/pubmed/21265847. 215

J. M. Yashari, C. G. Duncan, and F. M. Duerr. Evaluation of a novel canine activity monitor for at-home physical activity analysis. *BMC Veterinary Research*, 11(146), 2015. ISSN 1746-6148. doi: 10.1186/s12917-015-0457-y. URL http://www.biomedcentral.com/1746-6148/11/146. 255

X. Zhou and M. Stephens. Genome-wide Efficient Mixed Model Analysis for Association Studies. *Nature genetics*, 44(7):821–824, 2012. doi: 10.1038/ng.2310. Genome-wide. URL http://www.nature.com/ng/journal/v44/n7/abs/ng.2310.html. 221